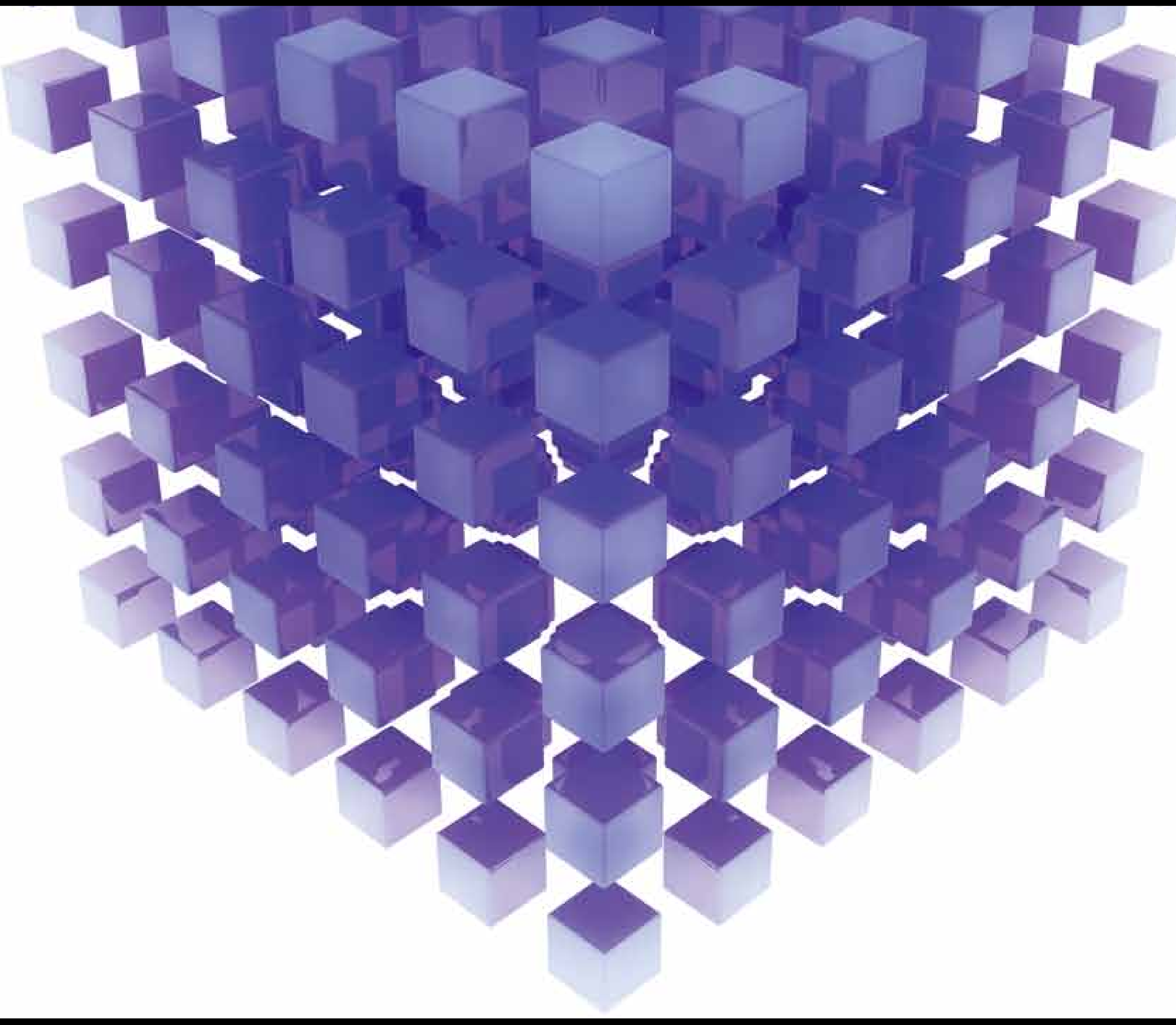
Mathematical Problems in Engineering

Multidisciplinary Design Optimization in Engineering

Guest Editors: Zhuming Bi, Lihui Wang, Chong Wu, Guilin Yang, and Dan Zhang



Multidisciplinary Design Optimization in Engineering

Mathematical Problems in Engineering

Multidisciplinary Design Optimization in Engineering

Guest Editors: Zhuming Bi, Lihui Wang, Chong Wu,
Guilin Yang, and Dan Zhang

Editorial Board

Contents

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 351097, 2 pages
http://dx.doi.org/10.1155/2013/351097

Editorial

Multidisciplinary Design Optimization in Engineering

Zhuming Bi,[1] Lihui Wang,[2] Chong Wu,[3] Guilin Yang,[4] and Dan Zhang[5]

[1] *Department of Engineering, Indiana University-Purdue University Fort Wayne, 2101 East Coliseum Boulevard, Fort Wayne, IN 46805-149, USA*
[2] *Department of Production Engineering, KTH Royal Institute of Technology, Brinellvägen 68, 100 44 Stockholm, Sweden*
[3] *School of Management, Harbin Institute of Technology, Harbin 150001, China*
[4] *Mechatronics Group, Singapore Institute of Manufacturing Technology, 71 Nanyang Drive, Singapore 638075*
[5] *Faculty of Engineering and Applied Science, University of Ontario Institute of Technology, 2000 Simcoe Street, ON, Canada L1H 7K4*

Correspondence should be addressed to Zhuming Bi; biz@ipfw.edu

Received 20 August 2013; Accepted 20 August 2013

System complexity depends on the number of design parameters in a system as well as the dynamic characteristics of design parameters in the time domain. Complex systems are generally featured as a large number of design parameters, residential dynamic environments, strongly coupled system behaviors, and the needs of multiple design criteria. Modern systems are becoming more and more complex due to numerous factors such as the involvement of multidisciplinary behaviors, increasing functionalities and components, growing amount of information associated with system interactions, and the uncertainties and changes involved in systems.

Multidisciplinary design optimization (MDO) is essential to the design and operation of a complex system. MDO takes into consideration all relevant disciplines simultaneously to find the global optimum which is superior to a solution from a sequence of local optimizations in individual disciplines. MDO has been successfully applied in the designs of many complex systems such as aircrafts, automobiles, shipbuilding, computers, and civic infrastructures. However, a simultaneous consideration of multiple disciplines increases the complexity of design problems, and the products or systems in future tend to be even complicated. Continuing research and development efforts are in demand to advance the theories and methodologies of MDO. This special issue is dedicated to the recent progresses on theories, methodologies, and applications of MDO. The paper solicitation has received numerous responses from the researchers in the field, and, through a rigorous peer-review process, 13 papers have been accepted to represent the recent advancement of design theories, methodologies, and case studies related to MDO. These papers can be roughly classified into two catalogues: solving strategies and applications.

R. A. Borges et al. discussed the formulation and solution of a system optimization using heuristic methods. Three bioinspired algorithms (Bees Colony algorithm, Firefly Colony algorithm, and Fish Swarm algorithm) are applied and compared to maximize the suppression bandwidth of a nonlinear damped system. MDO involves multiple design criteria, and two alternative ways to deal with the confliction of objectives in linear programming are Pareto set and Data Envelopment Analysis (DEA). M. Li provided a new representation of the Pareto set, which is generic to optimization problems with multi-design criteria. This representation was applied in Data Envelopment Analysis (DEA) models which simplified the derivation and acquisition of the properties associated with the Pareto set. J. W. Wang et al. presented a new algorithm with an integration of the Particle Swarm Optimization and predatory search strategy. The new algorithm was compared with others and showed its advantages in tackling the balance of exploitation and exploration in optimization. The heuristic algorithms can generally deal with any optimization problems; however, they have the limitations of inconsistency and low convergence rate; in particular, it is challenging to be applied in an optimization with expensive fitness functions. A surrogate model can be integrated to alleviate these limitations. T. Kunakote and S. Bureerat proposed to integrate the Pareto evolutionary algorithm with several surrogate models. Different integration scenarios were considered for eight structural design problems. Computation and time are the critical measures for the efficiency of optimization methodologies. Time relates to the number of iterations in the solving process. E. Devendorf and K. Lewis discussed the estimation of the number of iterations for distributed design processes; the game theory is integrated with the conventional discrete system theory for approximation.

B. Li et al. developed a methodology to design stiffened plate structures using evolutionary algorithms. X. Liang et al. investigated the effect of shock wave on a plate and associated fluid-structure interaction and thickness effects. An analytical approach was proposed for the transient analysis of a homogenous rectangular plate; the Mindlin plate theory and the Navier solution were integrated with Laplace inversion technique. W. Huang and S. K. Oh proposed the new fuzzy inference systems based on a multiobjective opposition-based space search algorithm (MOSSA); the developed systems are used for system identification to enhance the flexibility of fuzzy models. The MOSSA is a space search algorithm with an opposition-based learning method; this new method employs an "opposite numbers" mechanism to accelerate the convergence in optimization. T. Yifei et al. formulated an energy model of bridge cranes, and the Finite Element Method was used for the structure-level optimization. Time dependence is an important issue in the design optimization. K. Meruva et al. tackled the formulation of optimization with a consideration of the fatigue life, and solved the design problem with an integration of analytical models and numerical approaches. This work has its significance at both aspects of design and optimization of components and large scale complex system. C. Liu et al. designed a new variable reluctance (VR) resolver for the angle measurement. The models for the structure and output voltages of signal windings are formulated, and Finite Element Analysis (FEA) was applied to predict the machine performance. In optimizing a planetary gear system, Q. Huang et al. proposed a dynamic model, which took into account time-varying mesh stiffness, excitation fluctuation, and gear backlash nonlinearities. The developed nonlinear differential equations were solved by variable step-size Runge-Kutta algorithm, and an optimization model was developed to minimize the system vibration. V. V. Kumar et al. discussed the support of the optimization involved in design of fixtures and product lines; they presented a web-based framework to perform fixture design and assembly line production concurrently. The framework was capable of dealing with the spatial and generational varieties, and computer-assisted tools were embedded to select suppliers and optimize fixtures.

We wish the collected works are helpful to the readers in the field of MDO. Due to the time constraint, we are not able to include more papers covering other important issues such as the formulation of complex optimization systems and the MDO applications in large-scale systems. Readers are encouraged to submit their new contributions to this journal.

Acknowledgments

The guest editorial team is grateful to the authors of the special issue for their contributions and to the reviewers for their valuable comments on the submissions.

Zhuming Bi
Lihui Wang
Chong Wu
Guilin Yang
Dan Zhang

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 737502, 12 pages
http://dx.doi.org/10.1155/2013/737502

Research Article

Application of Three Bioinspired Optimization Methods for the Design of a Nonlinear Mechanical System

Romes Antonio Borges,[1] Fran Sérgio Lobato,[2] and Valder Steffen Jr.[3]

[1] *Department of Mathematics, Federal University of Goiás, Avendia Dr. Lamartine Pinto de Avelar 1120, 75704-220 Catalão, GO, Brazil*
[2] *School of Chemical Engineering, Federal University of Uberlândia, Avendia João Naves de Ávila 2121, Campus Santa Mônica, P.O. Box 593, 38408-144 Uberlândia, MG, Brazil*
[3] *School of Mechanical Engineering, Federal University of Uberlândia, Avendia João Naves de Ávila 2121, Campus Santa Mônica, P.O. Box 593, 38408-144 Uberlândia, MG, Brazil*

Correspondence should be addressed to Valder Steffen Jr.; vsteffen@mecanica.ufu.br

Received 21 January 2013; Revised 9 June 2013; Accepted 10 June 2013

Academic Editor: Chong Wu

The present work focuses on the optimal design of nonlinear mechanical systems by using heuristic optimization methods. In this context, the nonlinear optimization problem is devoted to a two-degree-of-freedom nonlinear damped system, constituted of a primary mass attached to the ground by a linear spring and a secondary mass attached to the primary system by a nonlinear spring. This arrangement forms a nonlinear dynamic vibration absorber (nDVA), which is used in this contribution as a representative example of a nonlinear mechanical system. The sensitivity analysis of the suppression bandwidth, namely, the frequency range over which the ratio of the main mass displacement amplitude to the amplitude of the forcing function is less than unity, with respect to the design variables that characterize the nonlinear system based on the first order finite differences is presented. For illustration purposes the optimization problem is written as to maximize the suppression bandwidth by using three recent bioinspired optimization methods: Bees Colony Algorithm, Firefly Colony Algorithm, and Fish Swarm Algorithm. The results are compared with other evolutionary strategies.

1. Introduction

In various mechanical design contexts, engineers have to deal with nonlinear systems in which the dynamic response depends on a number of physical parameters. An example of such a system is the so-called dynamic vibration absorber (DVA). The DVA is used to reduce noise and vibration in various types of engineering systems, such as compressors, robots, ships, power lines, airplanes, and helicopters. Much of the knowledge available to date is compiled in the original patent by Frahm [1], in the books by Hartog [2] and Koronev and Reznikov [3] and in some review works such as the one by Steffen Jr and Rade [4]. In the last two decades, a great deal of effort has been devoted to the development of mathematical models for characterizing the mechanical behavior of nonlinear dynamic vibration absorbers (nDVAs) accounting for their typical dependence on design parameters that influence the nonlinear behavior of the system. A particular type of nDVA is the so-called viscoelastic neutralizer as studied by Espíndola and Bavastri [5]. Borges et al. [6] proposed and determined the robust optimal design of an nDVA combining sensitivity analysis and multiobjective optimization. Different techniques have been proposed to design viscoelastic vibration absorbers [7, 8]. Besides the well-known complexity of the modeling strategy involved in nonlinear dynamics, some general methodologies have been suggested and have been shown to be particularly suitable to be used in combination with structural systems discretization. This aspect makes them very attractive for the modeling of nonlinear dynamic vibration absorbers. Among these strategies, the theoretical study proposed by Nissen and coworkers [9] and Pai and Schulz [10] should be mentioned, in which techniques to improve the stability and efficiency of nDVAs into a frequency band of interest have been proposed, leading to refined nDVAs. Also, Rice and McCraith [11] and Shaw and coworkers [12] suggested optimization strategies to be applied to the design of nDVAs by applying an asymmetric nonlinear Duffing-type element incorporated in the suspension for narrow-band absorption applications.

Nowadays, different approaches based on optimization methods have been proposed to solve mechanical design problems. Cao and Wu [13] proposed a cellular automata based genetic algorithm applied to mechanical systems design. In this algorithm, the individuals in the population are mapped onto a cellular automata to realise the locality and neighbourhood. The mapping is based on the individuals' fitness and the Hamming distances between individuals. The selection of individuals is controlled based on the structure of cellular automata, to avoid the fast population diversity loss and improve the convergence performance during the genetic search. He and coworkers He, Prempain, and Wu [14] proposed an improved particle swarm optimizer (PSO) associated with the fly-back mechanism to maintain a feasible population. This optimization strategy was applied to mechanical systems design involving problem-specific constraints and mixed variables such as integer, discrete, and continuous variables. In this sense, biological systems have contributed significantly to the development of new optimization techniques. These methodologies, known as Bioinspired Optimization Methods (BiOMs), are based on strategies that seek to mimic the behavior observed in species found in the nature to update a population of candidates to solve optimization problems [15, 16]. These systems have the capacity to notice and modify their environment in order to seek for diversity and convergence. In addition, this capacity makes possible the communication among the agents (individuals of population) that capture the changes in the environment generated by local interactions [17].

In this context, nature-inspired algorithms have contributed significantly to the development of new optimization techniques. Among the most recent bioinspired strategies stand the Bees Colony Algorithm (BCA) [18], the Firefly Colony Algorithm (FCA) [19], and the Fish Swarm Algorithm (FSA) [20]. The BCA is based on the behavior of bees colonies in their search of raw materials for honey production. According to Lucic and Teodorovic [21], in each hive groups of bees (called scouts) are recruited to explore new areas in search for pollen and nectar. These bees, returning to the hive, share the acquired information so that new bees are indicated to explore the best regions visited in an amount proportional to the previously passed assessment. Thus, the most promising regions are best explored and eventually the worst end up being discarded. Every iteration, this cycle repeats itself with new areas being visited by scouts. The FCA mimics the patterns of short and rhythmic flashes emitted by fireflies in order to attract other individuals to their vicinities. The corresponding optimization algorithm is formulated by assuming that all fireflies are unisex, so that one firefly will be attracted to all other fireflies. Attractiveness is proportional to their brightness, and for any two fireflies, the less bright will attract (and thus move to) the brighter one. However, the brightness can decrease as their distance increases and if there are no fireflies brighter than a given firefly it will move randomly. The brightness is associated with the objective function for optimization purposes [19]. Finally, the FSA is a random search algorithm based on the behaviour of fish swarm observed in nature. This behavior can be summarized as follows [20]: *random behavior*—in general, fish looks at random for food and other companion; *searching behavior*—when the fish discovers a region with more food, it will go directly and quickly to that region; *swarming behavior*—when swimming, fish will swarm naturally in order to avoid danger; *chasing behavior*—when a fish in the swarm discovers food, the others will find the food dangling after it; *leaping behavior*—when fish stagnates in a region, a leap is required to look for food in other regions.

In this way, the present work is dedicated to presenting alternative techniques for the optimal design of a nonlinear dynamic vibration absorber (nDVA) by using heuristic optimization methods to maximize the suppression bandwidth. For this aim, this contribution focuses on the theoretical study and numerical simulation of a two-degree-of-freedom nonlinear damped system, constituted of a primary mass attached to the ground by a linear spring and a secondary mass attached to the primary system by a nonlinear spring (nDVA). A previous contribution regarding this type of mechanical system can be found in Borges et al. [6]. Then, the three previously mentioned optimization techniques are applied to the design of an nDVA for illustration purposes; however, they are intended to be general in the sense that they can be applied to design different types of nonlinear mechanical systems. This work is organized as follows. The mathematical formulation of the nonlinear dynamic system and sensitivity analysis are presented in Sections 2 and 3, respectively. A review of the BCA, the FCA, and FSA are presented in Section 4. The results and discussion are described in Section 5. Finally, the conclusions and suggestions for future work conclude the paper.

2. Mathematical Modeling of the Nonlinear Dynamic System

Consider the two-degree-of-freedom (d.o.f.) model shown in Figure 1.

This device consists of a damped primary system attached to the ground by a suspension that includes either a linear or a nonlinear spring, and a damped secondary mass coupled to the primary system by a spring with nonlinear characteristics. The external force applied to the primary system is given by the following expression:

$$F_1 = p\cos(\omega t). \tag{1}$$

The constitutive forces of the springs are given by

$$r_i(x_i) = K_i x_i + k_i^{\mathrm{nl}} x_i^3 \quad i = 1, 2, \tag{2}$$

where x_1 represents the displacement of the primary system with respect to the ground, and x_2 is the displacement of

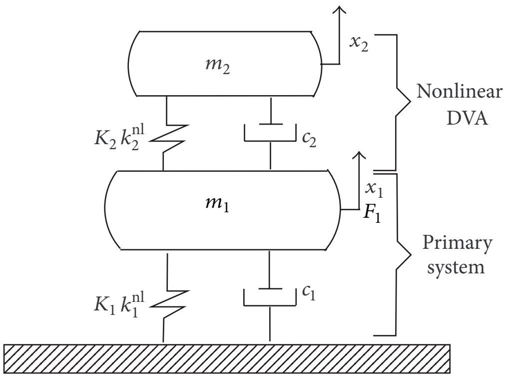


Figure 1: Two degree-of-freedom damped system [6].

the DVA with respect to the primary system. In the model above, the dampers are linear, however springs have nonlinear characteristics, where K_i and k_i^{nl} represent, respectively, the linear and nonlinear coefficients of the springs.

With the aim of obtaining the dimensionless normalized equations of motion for the nonlinear system, the displacements are normalized with respect to the length of a given vector x_c [22] according to the following expression:

$$y_i = \frac{x_i}{x_c}. \quad (3)$$

In addition, one introduces the following relations to the system:

$$\begin{gathered}
\overline{\omega}^2 = \frac{k_i}{m_i}, \qquad \zeta_i = \frac{c_i}{2\sqrt{k_i m_i}}, \\
\delta_i = 2\zeta_i\omega_i, \qquad \eta_i = \omega_i^2, \qquad \varepsilon_i = \frac{k_i^{nl} x_c^2}{m_i \omega^2}, \\
\rho = \frac{\omega_2}{\omega_1}, \qquad P = \frac{p}{m_1 \overline{\omega}^2 x_c}, \\
\Omega = \frac{\omega}{\overline{\omega}_1}, \qquad \mu = \frac{m_2}{m_1}.
\end{gathered} \quad (4)$$

By applying Newton's second law, and after some algebraic manipulations, the following normalized equation of motion of the nonlinear dynamic system can be expressed under the following matrix form:

$$\mathbf{M}\ddot{\mathbf{y}}(t) + \mathbf{C}\dot{\mathbf{y}}(t) + \mathbf{K}\mathbf{y}(t) = \mathbf{f}(t), \quad (5)$$

where the normalized mass, damping and stiffness matrices are expressed, respectively, by the following relations:

$$\mathbf{M} = \begin{bmatrix} 1+\mu & \mu \\ \mu & \mu \end{bmatrix}, \qquad \mathbf{C} = \begin{bmatrix} \delta_1 & 0 \\ 0 & \mu\delta_2 \end{bmatrix},$$
$$\mathbf{K} = \begin{bmatrix} \eta_1 & 0 \\ 0 & \eta_2 \end{bmatrix}. \quad (6)$$

The normalized displacement and force vectors are given by:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \qquad \mathbf{f} = \begin{bmatrix} P\cos(\tau)\,\varepsilon_1 y_1^3 \\ -\varepsilon_2 y_2^3 \mu \end{bmatrix}. \quad (7)$$

It is important to emphasized that the present contribution is dedicated to maximize the suppression bandwidth.

2.1. Steady-State Response of the Nonlinear System. In this paper, the Krylov-Bogoliubov method [23] is used to integrate the matrix equation of motion (1). This method leads to an approximate solution of nonlinear differential equations. The process is based on the following transformation of variables:

$$\mathbf{y} = \mathbf{u}(\tau)\cos(\tau) + \mathbf{v}(\tau)\sin(\tau), \quad (8)$$

where $\tau = \omega t$ is the time dependence of $\mathbf{u} = (u_1 u_2)^T$, and $\mathbf{v} = (v_1 v_2)^T$ is assumed to be small for high-order terms, such as the vectors $\mathbf{u}$ and $\mathbf{v}$.

After mathematical manipulation, a nonlinear algebraic system with four equations and four variables u_1, u_2, v_1 and v_2 is obtained as follows:

$$\begin{gathered}
\left(1+\mu-\omega_1^2\right)u_1 + \mu u_2 - 2\zeta_1\omega_1 v_1 \\
- \frac{3\varepsilon_1\left(u_1^2+v_1^2\right)u_1}{4} + \beta\omega_1^2 = 0, \\
\mu u_1 + \left(\mu - \mu\rho^2\omega_1^2\right)u_2 - \mu\left(2\zeta_2\rho\omega_1^2 v_2 + \frac{3\varepsilon_2\left(u_2^2+v_2^2\right)u_2}{4}\right) \\
\left(\omega_1^2 - 1 - \mu\right)v_1 - \mu v_2 - 2\zeta_1\omega_1 u_1 + \frac{3\varepsilon_1\left(u_1^2+v_1^2\right)v_1}{4} = 0, \\
\mu v_1 + \left(\mu\rho^2\omega_1^2 - \mu\right)v_2 \\
- \mu\left(2\zeta_2\rho\omega_1^2 u_2 - \frac{3\varepsilon_2\left(u_2^2+v_2^2\right)v_2}{4}\right) = 0.
\end{gathered} \quad (9)$$

The system represented by (9) should be numerically solved. Then, the values of u_1, u_2, v_1 and v_2 can be calculated and the vibration amplitudes of the primary and secondary masses of the nonlinear DVA are obtained. The amplitude values are given by X_1 and X_2 according to the following equation:

$$X_i = \left(u_i^2 + v_i^2\right)^{0.5} \quad i = 1, 2 \quad (10)$$

3. Sensitivity Analysis of Structural Responses

In a mechanical system, the parameters of mass, stiffness, and damping establish the dependence with respect to a set of design parameters, which includes physical and geometrical characteristics and the parameters that are associated with the nonlinearities [24]. Such functional dependence can be expressed in a general form as follows:

$$\mathbf{r} = \mathbf{r}(\mathbf{M}(\mathbf{p}), \mathbf{C}(\mathbf{p}), \mathbf{K}(\mathbf{p})), \quad (11)$$

where $\mathbf{r}$ and $\mathbf{p}$ designate vectors of structural responses and design parameters, respectively.

(1) Initialize population with random solutions
(2) Evaluate fitness of the population
(3) While (stopping criterion not met)
(4) Select sites for neighborhood search
(5) Recruit bees for selected sites and evaluate fitnesses
(6) Select the fittest bee from each site
(7) Assign remaining bees to search randomly and evaluate their fitnesses
(8) End while

Algorithm 1: Basic step in the Bees Colony Algorithm [18].

The sensitivity of the structural responses with respect to a given parameter p_i evaluated for a given set of values of the design parameter p^0 is defined as the following partial derivative:

$$\frac{\partial \mathbf{r}}{\partial \mathbf{p}_i} = \lim_{\Delta \mathbf{p}_i \to 0} \left[R_1 + R_2 \right], \tag{12}$$

where

$$R_1 = \frac{\mathbf{r}\left(\mathbf{M}\left(\mathbf{p}_i^0 + \Delta\mathbf{p}_i\right), \mathbf{C}\left(\mathbf{p}_i^0 + \Delta\mathbf{p}_i\right), \mathbf{K}\left(\mathbf{p}_i^0 + \Delta\mathbf{p}_i\right)\right)}{\Delta\mathbf{p}_i},$$
$$R_2 = \frac{\mathbf{r}\left(\mathbf{M}\left(\mathbf{p}_i^0\right), \mathbf{C}\left(\mathbf{p}_i^0\right), \mathbf{K}\left(\mathbf{p}_i^0\right)\right)}{\Delta\mathbf{p}_i}, \tag{13}$$

where $\Delta\mathbf{p}_i$ is an arbitrary variation applied to the current value of parameter $\mathbf{p}_i^0$, while all other parameters remain unchanged. The sensitivity with respect to $\mathbf{p}_i$ can be estimated by finite differences by computing successively the responses corresponding to $\mathbf{p}_i = \mathbf{p}_i^0$ and $\mathbf{p}_i = \mathbf{p}_i^0 + \Delta\mathbf{p}_i$. Such procedure is an estimated approach enabling to calculate the sensitivity of the dynamic system responses with respect to small modifications introduced in the design parameters. Moreover, the results depend upon the choice of the value of the parameter increment $\Delta\mathbf{p}_i$. Another strategy consists in computing the analytical derivatives, if possible, of the structural responses with respect to the parameters of interest. However, this approach is not considered in the present paper.

4. Bioinspired Algorithms

Algorithms based on swarm intelligence principles are common in the literature due to the ability to find global solution in mono- and multiobjective contexts, different from algorithms based on gradients. According to Andrés and Lozano [25], swarm-based intelligence is artificial intelligence technique based on the study of collective behaviour in self-organizing systems, composed of a population of individuals, which takes effect between each other and environment. Although these systems do not have any central control of the individual behaviour, interaction between individuals and simple behaviour between them usually lead to detection of aggregate behaviour, which is typical for whole colony. This could by observed by ants, bees, birds, or bacteria in the nature. By inspiration of these colonies were develop algorithms called swarm-based intelligence and are successfully applied for solving complicated optimization problems [26]. In this contribution, three recent bioinspired optimization methods, Bees Colony Algorithm, Firefly Colony Algorithm, and Fish Swarm Algorithm, are considered as optimization strategies.

4.1. Bee Colony Algorithm. This optimization algorithm is based on the behavior of a colony of honey bees. The colony can extend itself over long distances and in multiple directions simultaneously to exploit a large number of food sources. In addition, the colony of honey bees presents as characteristic the capacity of memorization, learning, and transmission of information, thus forming the so-called swarm intelligence [27].

In a colony, the foraging process begins by scout bees being sent to search randomly for promising flower patches. When they return to the hive, those scout bees that found a patch which is rated above a certain quality threshold (measured as a combination of some constituents, such as sugar content) deposit their nectar or pollen and go to the waggle dance.

This dance is responsible for the transmission (colony communication) of information regarding a particular flower patch: the direction in which it will be found, its distance from the hive, and its quality rating (or fitness) [27]. The waggle dance enables the colony to evaluate the relative merit of different paths according to both the quality of the food they provide and the amount of energy needed to harvest it [28]. After waggle dancing, the dancer (scout bee) goes back to the flower patch with follower bees that were waiting inside the hive. More follower bees are sent to more promising patches. This allows the colony to gather food quickly and efficiently. While harvesting from a patch, the bees monitor its food level. This is necessary to decide upon the next waggle dance when they return to the hive [28]. If the patch is still good enough as a food source, then it will be advertised in the waggle dance, and more bees will be recruited to that source. In this context, Pham et al. [18] proposed an optimization algorithm inspired by the natural foraging behavior of honey bees (Bees Colony Algorithm (BCA)) as presented in Algorithm 1.

The BCA requires a number of parameters to be set, namely, the number of scout bees (n), number of sites selected for neighborhood search (out of n visited sites) (m), number of top-rated (elite) sites among m selected sites (e), number of bees recruited for the best e sites (nep), number of bees recruited for the other $(m - e)$ selected sites, neighborhood search (ngh), and the stopping criterion.

The BCA starts with the n scout bees being placed randomly in the search space. The fitnesses of the sites visited by the scout bees are evaluated in step 2.

In step 4, bees that have the highest fitnesses are chosen as selected bees, and sites visited by them are chosen for neighborhood search. Then, in steps 5 and 6, the algorithm conducts searches in the neighborhood of the selected sites, assigning more bees to search near to the best e sites. The bees can be chosen directly according to the fitnesses associated with the sites they are visiting.

Alternatively, the fitness values are used to determine the probability of the bees being selected. Searches in the neighborhood of the best e sites, which represent more promising solutions, are made more detailed by recruiting more bees to follow them than the other selected bees. Together with scouting, this differential recruitment is a key operation of the BCA.

However, in step 6, for each patch only the bee with the highest fitness will be selected to form the next bee population. In nature, there is no such restriction. This restriction is introduced here to reduce the number of points to be explored. In step 7, the remaining bees in the population are assigned randomly around the search space scouting for new potential solutions.

In the literature, various applications using this bioinspired approach can be found, such as modeling combinatorial optimization transportation engineering problems [21], engineering system design [15, 29], mathematical function optimization [18], transport problems [30], dynamic optimization [31], optimal control problems [32], and parameter estimation in control problems [16, 33] (http://mf.erciyes.edu.tr/abc/).

4.2. Firefly Colony Algorithm. The FCA is based on the characteristic of the bioluminescence of fireflies, insects notorious for their light emission. According to Yang [19], biology does not have a complete knowledge to determine all the utilities that firefly luminescence can bring to, but at least three functions have been identified: (i) as a communication tool and appeal to potential partners in reproduction, (ii) as a bait to lure potential prey for the firefly, and (iii) as a warning mechanism for potential predators reminding them that fireflies have a bitter taste.

In this way, the bioluminescent signals are known to serve as elements of courtship rituals (in most cases, the females are attracted by the light emitted by the males), methods of prey attraction, social orientation, or as a warning signal to predators [34].

It was idealized some of the flashing characteristics of fireflies so as to develop firefly-inspired algorithms. For simplicity the following three idealized rules are used [35]:

(1) all fireflies are unisex so that one firefly will be attracted to other fireflies regardless of their sex;

(2) attractiveness is proportional to their brightness; thus, for any two flashing fireflies, the less brighter will move towards the brighter one. The attractiveness is proportional to the brightness and they both decrease as their distance increases. If there is no brighter one than a particular firefly, it will move randomly;

(3) the brightness of a firefly is affected or determined by the landscape of the objective function. For a maximization problem, the brightness can simply be proportional to the value of the objective function.

According to Yang [19], in the firefly algorithm there are two important issues: the variation of light intensity and the formulation of the attractiveness. For simplicity, it is always assumed that the attractiveness of a firefly is determined by its brightness, which in turn is associated with the encoded objective function.

This swarm intelligence optimization technique is based on the assumption that the solution of an optimization problem can be perceived as agent (firefly) which glows proportionally to its quality in a considered problem setting. Consequently, each brighter firefly attracts its partners (regardless of their sex), which makes the search space being explored more efficiently. The algorithm makes use of a synergic local search. Each member of the swarm explores the problem space taking into account results obtained by others, still applying its own randomized moves as well. The influence of other solutions is controlled by the attractiveness value [34].

According to Lukasik and Zak [34], the FCA is presented in the following. Consider a continuous constrained optimization problem where the task is to minimize the cost function $f(x)$ as follows

$$f\left(x^*\right) = \min_{x \in S} f\left(x\right). \tag{14}$$

Assume that there exists a swarm of m_a agents (fireflies) solving the above mentioned problem iteratively, and x_i represents a solution for a firefly i in algorithm's iteration k, whereas $f(x_i)$ denotes its cost. Initially, all fireflies are dislocated in S (randomly or employing some deterministic strategy). Each firefly has its distinctive attractiveness λ which implies how strong it attracts other members of the swarm. As the firefly attractiveness, one should select any monotonically decreasing function of the distance $r_j = d(x_i, x_j)$ to the chosen firefly j, for example, the exponential function

$$\lambda = \lambda_0 \exp\left(-\gamma r_j\right), \tag{15}$$

where λ_0 and γ are predetermined algorithm parameters: maximum attractiveness value and absorption coefficient, respectively. Furthermore, every member of the swarm is characterized by its light intensity I_i which can be directly expressed as an inverse of a cost function $f(x_i)$. To effectively explore considered search space S, it is assumed that each firefly i is changing its position iteratively taking into account two factors: attractiveness of other swarm members with higher light intensity, for example, $I_j > I_i$, for all $j = 1, \ldots, m_a$, $j \neq i$ which is varying across distance, and a fixed random step vector u_i. It should be noted as well that if no brighter firefly can be found, only such randomized step is being used.

Thus, moving at a given time step t of a firefly i toward a better firefly j is defined as

$$x_i^t = x_i^{t-1} + \lambda \left(x_j^{t-1} - x_i^{t-1} \right) + \alpha \left(\text{rand} - 0.5 \right), \quad (16)$$

where the second term on the right side of the equation inserts the factor of attractiveness, λ while the third term, governed by α parameter, governs the insertion of certain randomness in the path followed by the firefly, rand is a random number between 0 and 1.

In the literature, few works using the FCA can be found. In this context, is emphasized application in parameter estimation in control problems [16], continuous constrained optimization task [34], multimodal optimization [36], solution of singular optimal control problems [37], and economic emissions load dispatch problem [38].

4.3. Fish Swarm Algorithm. In the development of the FSA, based on fish swarm and observed in nature, the following characteristics are considered [20, 39]: (i) each fish represents a candidate solution of optimization problem; (ii) food density is related to an objective function to be optimized (in an optimization problem, the amount of food in a region is inversely proportional to the objective function value); (iii) the aquarium is the design space where the fish can be found.

As noted earlier, the fish weight at the swarm represents the accumulation of food (e.g., the objective function) received during the evolutionary process. In this case, the weight is an indicator of success [20, 39].

Basically, the FSA presents four operators classified into two classes: "food search" and "movement." Details on each of these operators are shown in the following.

4.3.1. Individual Movement Operator. This operator contributes to the movement (individual and collective) of fishes in the swarm. Each fish updates its position by using

$$x_i (t+1) = x_i (t) + \text{rand} \times s_{\text{ind}}, \quad (17)$$

where x_i is the final position of fish i at current generation, rand is a random generator, and s_{ind} is a weighted parameter.

4.3.2. Food Operator. The weight of each fish is a metaphor used to measure the success of food search. The higher the weight of a fish, the more likely this fish will be in a potentially interesting region in the design space.

According to Madeiro [39], the amount of food that a fish eats depends on the improvement in its objective function in the current generation and the greatest value considering the swarm. The weight is updated according to

$$W_i (t+1) = W_i (t) + \frac{\Delta f_i}{\max (\Delta f)}, \quad (18)$$

where $W_i(t)$ is the fish weight i at generation t, and Δf_i is the difference found in the objective function between the current position and the new position of fish i. It is important to emphasize that $\Delta f_i = 0$ for the fish in the same position.

4.3.3. Instinctive Collective Movement Operator. This operator is important for the individual movement of fish when $\Delta f_i \neq 0$. Thus, only the fish whose individual execution of the movement resulted in improvement of their fitness will influence the direction of motion of the swarm, resulting in instinctive collective movement. In this case, the resulting direction (I_d), calculated using the contribution of the directions taken by the fish, and the new position of the ith fish are given by

$$I_d (t) = \frac{\sum_{i=1}^{N} \Delta x_i \Delta f_i}{\sum_{i=1}^{N} \Delta f_i},$$
$$x_i (t+1) = x_i (t) + I_d (t). \quad (19)$$

It is worth mentioning that in the application of this operator, the direction chosen by a fish that located the largest amount of food exerts the greatest influence on the swarm. Therefore, the instinctive collective movement operator tends to guide the swarm in the direction of motion chosen by the fish who found the largest amount of food in its individual movement.

4.3.4. Noninstinctive Collective Movement Operator. As noted earlier, the fish weight is a good indication of success in the search for food. In this way, the swarm weight is increasing, this means that the search process is successful. So, the "radius" of the swarm should decrease to other regions to be explored. Otherwise, if the swarm weight remains constant, the radius should increase to allow the exploration of new regions.

For the swarm contraction, the centroid concept is used. This is obtained by means of an average position of all fish weighted with the respective fish weights, according to

$$B (t) = \frac{\sum_{i=1}^{N} x_i W_i (t)}{\sum_{i=1}^{N} W_i (t)}. \quad (20)$$

If the swarm weight remains constant in the current iteration, all fish update their positions by using

$$x (t+1) = x (t) - s_{\text{vol}} \times \text{rand} \times \frac{x (t) - B (t)}{d (x (t), B (t))}, \quad (21)$$

where d is a function that calculates the Euclidean distance between the centroid and the current position of fish, and s_{vol} is the step size used to control fish displacements.

In the literature, few works using the FSA can be found. In this context, parameter estimation in control problems [16], feed forward neural networks [40], parameter estimation in engineering systems [41], combinatorial optimization problem [42], augmented Lagrangian fish swarm based method for global optimization [43], forecasting stock indices using radial basis function neural networks optimized [44], and hybridization of the FSA with the Particle Swarm Algorithm to solve engineering systems are examples of successful applications [45].

TABLE 1: Nominal values of design variables.

Parameters	Nominal values
ε_1	0.001
ε_2	0.01
β	0.1
ζ_1	0.01
ζ_2	0.01
μ	0.05
ρ	1.0

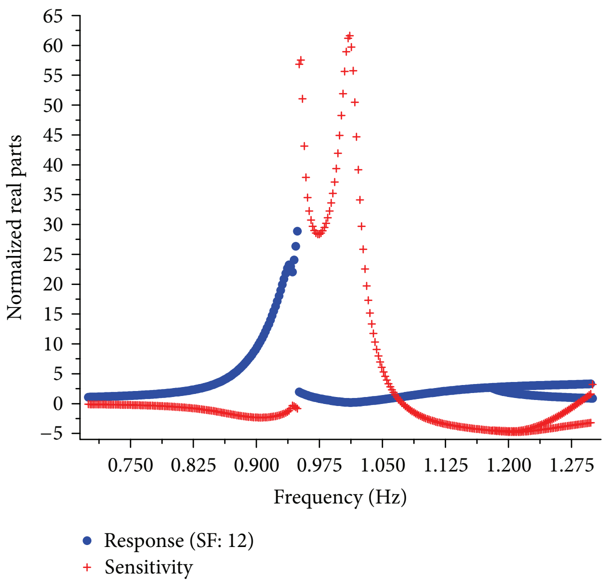


FIGURE 2: Sensitivity of $\mathbf{H}(\omega, \mathbf{p})$ with respect to ρ.

5. Results and Discussion

The following numerical example, studied in [6], is presented to illustrate the application of the proposed methodology to obtain the optimal design of a nDVA. As previously mentioned, the nDVA is used in the present contribution to represent a large class of nonlinear mechanical systems. Figure 1 depicts the test structure consisting of a primary mass attached to the ground by a nonlinear spring and coupled with an nDVA. The nominal values of the design parameters used to generate the dynamic responses of the nonlinear system are illustrated on Table 1. The computations performed consist in obtaining the driving point frequency responses associated to the displacement x_1.

5.1. Sensitivities of the Frequency Response with respect to Structural Parameters. To illustrate the computation procedure for the sensitivity of dynamic responses, numerical tests were performed by using the system configuration illustrated in Figure 1. As previously mentioned, the computations are devoted to obtaining the sensitivities of the driving point frequency responses, which are given by the elements of $\mathbf{H}(\omega, \mathbf{p})$.

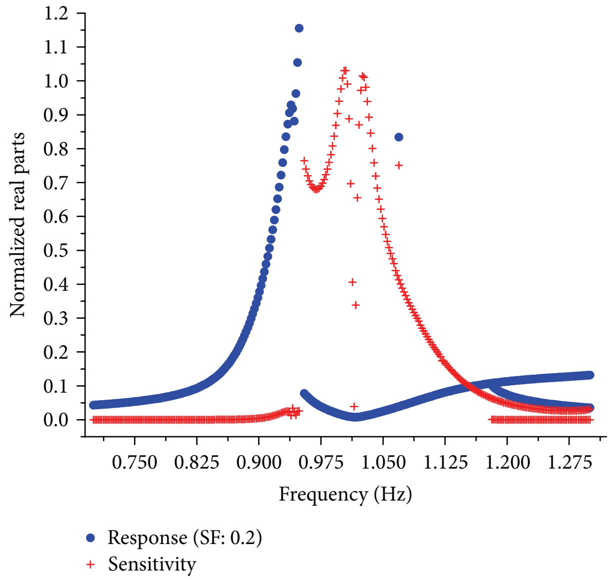


FIGURE 3: Sensitivity of $\mathbf{H}(\omega, \mathbf{p})$ with respect to ε_2.

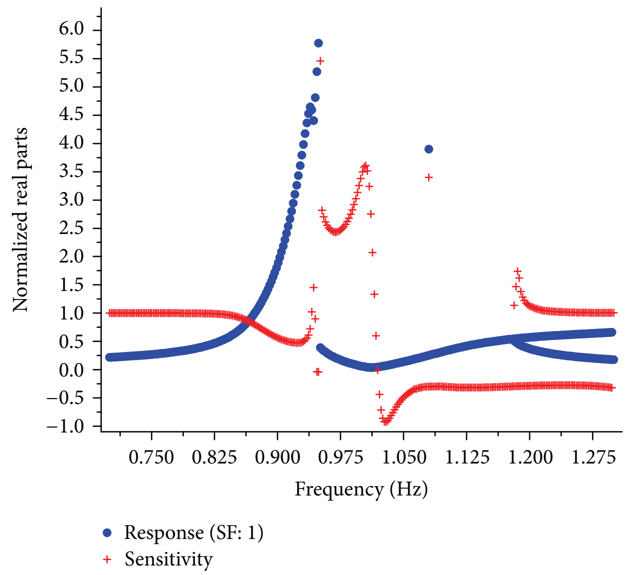


FIGURE 4: Sensitivity of $\mathbf{H}(\omega, \mathbf{p})$ with respect to β.

In this example, the normalized structural parameters ζ_1, ζ_2, ε_1, ε_2, β, μ, and ρ are considered as the design variables in the computation of the normalized sensitivities of the frequency responses with respect to a given parameter $\mathbf{p}$, $S_H^N(\omega, \mathbf{p})$. The normalized real parts of the approximated complex sensitivity functions calculated by finite differences (according to (17)) are shown in Figures 2, 3, 4, and 5, for which a variation of 20% of the nominal values of each design parameter was adopted. Also, in the same figures, the real parts of the frequency responses $\mathbf{H}(\omega, \mathbf{p})$, multiplied by

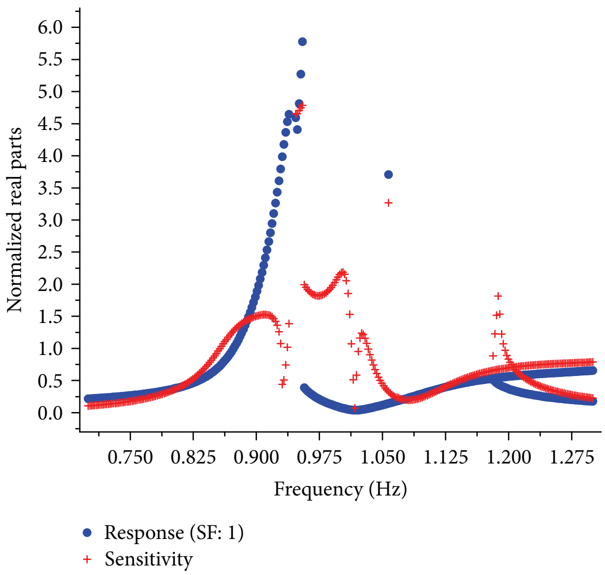


FIGURE 5: Sensitivity of $\mathbf{H}(\omega, \mathbf{p})$ with respect to μ.

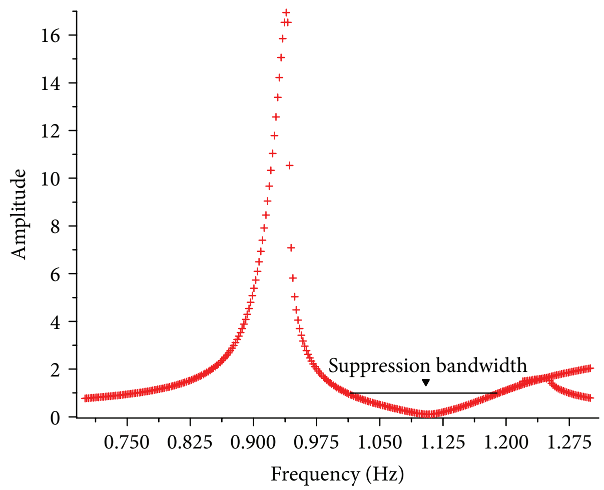


FIGURE 6: Representation of the objective function (maximize the suppression bandwidth).

convenient scale factors (SF), are shown. The sensitivity functions, denoted by $S_H^N(\omega, \mathbf{p})$, have been normalized according to the following scheme:

$$S_H^N(\omega, \mathbf{p}) = \left.\frac{\partial S_H(\omega, \mathbf{p})}{\partial \mathbf{p}}\right|_{(\omega, \mathbf{p}_0)} \frac{\mathbf{p}_0}{\mathbf{H}(\omega, \mathbf{p})}. \tag{22}$$

Based on the amplitudes and signs of the sensitivity functions, one can evaluate the degree of influence of the design variables upon the suppression bandwidth, in the frequency band of interest. In addition, the sensitivity analysis enables to decide among the design parameters those that will be retained in the optimization process. The parameters ζ_1, ζ_2, and ε_1 do not have a significant influence on the evaluation of the suppression bandwidth [6]. Consequently, these parameters are not considered as design variables in the optimization run. However, as shown in Figures 2, 3, 4, and 5, the degrees of influence of the parameters ε_2, μ, and ρ on the suppression bandwidth are significant and will be considered as design variables in the optimization process.

After having verified the influence of each design variable on the dynamic response of the nonlinear system, the interest now is to maximize the suppression bandwidth, as illustrated in Figure 6, using the bioinspired algorithms.

In these simulations, the following ranges are considered for the design parameters: $0.9 \leq \rho \leq 1.2$, $0.04 \leq \mu \leq 0.06$, $0.09 \leq \beta \leq 1.2$, and $0.009 \leq \varepsilon_2 \leq 0.012$.

In order to evaluate the performance of the three techniques proposed above (BCA, FCA, and FSA), the following parameters were used in the algorithms:

(1) BCA parameters: number of scout bees (10), number of bees recruited for the best e sites (5), number of bees recruited for the other selected sites (5), number of sites selected for neighborhood search (5), number of top-rated (elite) sites among m selected sites (5), neighborhood search (ngh) [10^{-3} 10^{-4} 10^{-6}], and generation number (50);

(2) FCA parameters: number of fireflies (15), maximum attractiveness value (0.9), absorption coefficient [0.7 0.9 1.0] and generation number (50);

(3) FSA parameters: number of fish (15), weighted parameter value (1), control fish displacements [10^{-1} 10^{-2} 10^{-3}], and generation number (50).

In order to examine the quality of the solution methodology considered, the Genetic Algorithm (GA) and the Particle Swarm Optimization (PSO) have been performed. The parameters used by GA and PSO are as follows:

(1) GA parameters: population size (15), crossover rate (0.8), mutation rate (0.01), and generation number (50);

(2) PSO parameters: population size (15), inertia weight (1), cognitive and social parameters (0.5), constriction factor (0.8), and generation number (50).

The stopping criterion used was the maximum number of iterations. Each case study was computed 20 times before calculating the average values. It should be emphasized that 1510 objective function evaluations for each algorithm are necessary.

In Table 2 the results (best (B), average (A) and worst (W)) obtained for the design of the nonlinear vibration absorber are presented.

This table shows that the three algorithms presented good estimates for the unknown parameters when compared with the GA and PSO. When the BCA is analyzed in terms of the neighborhood search parameter, the best result is obtained by using 10^{-3}, for example, a search region with smaller distances to exploit a large number of food sources. When the FCA is analyzed in terms of the absorption coefficient, the best result is found by using $\gamma = 1.0$, for example, thus emphasizing the local search. Finally, when the FSA is analyzed in terms of the control fish displacements, the best result is found by using $s_{\mathrm{vol}} = 10^{-1}$.

Table 2: Results obtained by BCA, FCA, FSA, GA, and PSO to design the nonlinear vibration absorber (OF = objective function).

			ρ	μ	β	ε_2	OF
BCA	10^{-3}	B	1.165	0.053	0.110	0.017	0.236
		A	1.189	0.055	0.102	0.015	0.230
		W	1.027	0.058	0.091	0.011	0.218
	10^{-4}	B	1.099	0.056	0.103	0.010	0.230
		A	1.134	0.058	0.095	0.011	0.224
		W	0.916	0.045	0.103	0.011	0.224
	10^{-6}	B	1.134	0.058	0.095	0.011	0.224
		A	0.981	0.046	0.095	0.010	0.218
		W	0.961	0.050	0.105	0.011	0.218
FCA	0.7	B	1.158	0.054	0.110	0.017	0.230
		A	1.158	0.054	0.110	0.017	0.230
		W	1.157	0.053	0.110	0.017	0.228
	0.9	B	1.148	0.055	0.110	0.018	0.230
		A	1.148	0.055	0.110	0.018	0.230
		W	1.148	0.055	0.110	0.018	0.230
	1.0	B	1.156	0.056	0.111	0.017	0.232
		A	1.148	0.055	0.110	0.018	0.230
		W	1.151	0.057	0.109	0.015	0.214
FSA	10^{-1}	B	1.166	0.052	0.109	0.016	0.235
		A	1.191	0.050	0.101	0.016	0.232
		W	1.019	0.053	0.092	0.010	0.219
	10^{-2}	B	1.101	0.054	0.099	0.099	0.232
		A	1.135	0.060	0.094	0.011	0.225
		W	0.921	0.048	0.102	0.011	0.225
	10^{-3}	B	1.135	0.061	0.099	0.010	0.226
		A	0.993	0.049	0.096	0.010	0.219
		W	0.971	0.051	0.111	0.010	0.214
GA		B	1.101	0.054	0.099	0.098	0.233
		A	1.136	0.062	0.092	0.012	0.226
		W	0.932	0.044	0.110	0.012	0.228
PSO		B	1.101	0.053	0.099	0.099	0.233
		A	1.132	0.061	0.095	0.010	0.224
		W	0.918	0.049	0.103	0.011	0.225

6. Conclusions

In this work, the Bees Colony Algorithm, the Firefly Colony Algorithm, and the Fish Swarm Algorithm were proposed as alternative techniques to obtain the optimal design of a nonlinear mechanical system. For illustration purposes they were applied in the design of a nonlinear dynamic vibration absorber. The system nonlinearity was introduced in the springs that connect the primary mass to the ground and the absorber to the primary mass, respectively.

As observed in Table 2, the algorithms led to satisfactory results in terms of the effectiveness of the optimal nDVA configuration and the number of objective function evaluations, when compared with GA and PSO strategies. However, the results obtained by the algorithms need yet to be better analyzed, so that more definitive conclusions can be drawn, for example, new mechanisms for diversity exploration should be studied.

The choice of the design variables is based on previous knowledge regarding their sensitivities with respect to the suppression bandwidth. It is worth mentioning that these parameters are directly associated with the effectiveness of the nDVA.

In terms of the system resolution, the equations of motion of the nonlinear two d.o.f. system were numerically integrated by using the so-called average method that provides an approximate solution to nonlinear dynamic problems. The nonlinear algebraic equations obtained were solved numerically enabling to determine the roots of the nonlinear algebraic equations.

It is worth mentioning that the nonlinearity factor is an important parameter to be investigated during the design procedure of nDVAs, due to its contribution to the reduction of the vibration level. However, care must be taken with high values of nonlinearity because of the instabilities introduced in the system. This point motivates an important aspect regarding the proposed methodology: obtaining the optimal spring nonlinear coefficient that guarantees the best stable solution for a given system.

Finally, the optimization techniques used in this paper can be successfully applied in the design of a great number of nonlinear mechanical systems.

As further work, we intend to extend the algorithms to the multiobjective context and assess the sensitivity of the parameters with respect to the effectiveness of the solution.

Nomenclature

B: Weight matrix
BCA: Bees Colony Algorithm
BiOM: Bioinspired Optimization Methods
C: Damping matrix
c_i: Damping coefficients
d: Euclidean distance
DVAs: Dynamic vibration absorbers
e: Number of top-rated sites
f: Objective function
$\mathbf{f}$: Force vector
F_1: Force applied to the primary system
FCA: Firefly Colony Algorithm
FSA: Fish Swarm Algorithm
$\mathbf{H}(\omega, \mathbf{p})$: Sensitivities of the driving point frequency responses
I: Light intensity
I_d: Fishes movement direction
K: Stiffness matrix
K_i: Linear coefficients of the springs
k_i^{nl}: Nonlinear coefficients of the springs
M: Mass matrix
m: Number of sites selected for neighborhood search
m_a: Number of fireflies
m_1: Primary mass
m_2: Secondary mass
n: Number of scout bees
nep: Number of bees recruited for the best e sites
ngh: Neighborhood search
$\mathbf{p}$: Design parameters
P: Normalized reference value
$\mathbf{r}$: Vectors of structural responses
r_1: Constitutive forces of the springs
R_i: Auxiliary variables vector
r_j: Distance function
S: Design space
S_H^N: Sensitivity function
s_{ind}: Weighted parameter
s_{vol}: Control fish displacements
$\mathbf{u}$: Auxiliary variables vector
$\mathbf{v}$: Auxiliary variables vector
x_1: Displacement of the primary system with respect to the ground
x_2: Displacement of the DVA with respect to the primary system
x_i: Design variables
x_c: Displacement reference value
X_i: Vibration amplitudes
y_i: Normalized displacements
W: Fish weight
α: Insertion of randomness
β: Normalized force parameter
δ_i: Normalized reference value
η_i: Normalized reference value
λ: Attractiveness
λ_0: Maximal attractiveness
ω: Normalized reference value
Ω: Normalized frequency value
ξ_i: Normalized coefficient of the damping
μ: Normalized mass ratio
ρ: Normalized frequency ratio
τ: Dimensionless time
γ: Absorption coefficient
ε_i: Normalized nonlinear coefficient of the springs.

Acknowledgments

The authors acknowledge the financial support provided by FAPEMIG, CNPq, and FUNAPE/UFG. The third author acknowledges the financial support provided by FAPEMIG and CNPq (INCT-EIE). This work was presented at the 21st Brazilian Congress of Mechanical Engineering, October 24-28, 2011, Natal, RN, Brazil).

References

[1] H. Frahm, "Device for Damping Vibrations of Bodies," US Patent 989, 958, 1911.

[2] J. P. D. Hartog, *Mechanical Vibrations*, McGraw-Hill Book Company, 1934.

[3] B. G. Koronev and L. M. Reznikov, *Dynamic Vibration Absorbers—Theory and Technical Applications*, John Wiley & Sons, 1993.

[4] V. Steffen Jr. and D. A. Rade, "Dynamic vibration absorber," in *Encyclopedia of Vibration*, pp. 9–26, Academic Press, 2001.

[5] J. J. Espíndola and C. A. Bavastri, "Viscoelastic neutralisers in vibration abatement: a non-linear optimization approach," *Brazillian Journal of Mechanical Sciences*, vol. 19, no. 2, pp. 154–163, 1997.

[6] R. A. Borges, A. M. G. Lima, and V. Steffen Jr., "Robust optimal design of a nonlinear dynamic vibration absorber combining sensitivity analysis," *Shock and Vibration*, vol. 17, no. 4-5, pp. 507–520, 2010.

[7] J. J. Espíndola, C. A. Bavastri, and E. M. de Oliveira Lopes, "Design of optimum systems of viscoelastic vibration absorbers for a given material based on the fractional calculus model," *Journal of Vibration and Control*, vol. 14, no. 9-10, pp. 1607–1630, 2008.

[8] J. J. Espíndola, P. Pereira, C. A. Bavastri, and E. M. Lopes, "Design of optimum system of viscoelastic vibration absorbers with a frobenius norm objective function," *Journal of the*

Brazilian Society of Mechanical Sciences and Engineering, vol. 31, no. 3, pp. 210–219, 2009.

[9] J. C. Nissen, K. Popp, and B. Schmalhorst, "Optimization of a non-linear dynamic vibration absorber," *Journal of Sound and Vibration*, vol. 99, no. 1, pp. 149–154, 1985.

[10] P. F. Pai and M. J. Schulz, "A refined nonlinear vibration absorber," *International Journal of Mechanical Sciences*, vol. 42, no. 3, pp. 537–560, 2000.

[11] H. J. Rice and J. R. McCraith, "Practical non-linear vibration absorber design," *Journal of Sound and Vibration*, vol. 116, no. 3, pp. 545–559, 1987.

[12] J. Shaw, S. W. Shaw, and A. G. Haddow, "On the response of the nonlinear vibration absorber," *International Journal of Non-Linear Mechanics*, vol. 24, no. 4, pp. 281–293, 1989.

[13] Y. J. Cao and Q. H. Wu, "A cellular automata based genetic algorithm and its application in mechanical design optimisation," in *Proceedings of the International Conference on Control (UKACC '98)*, pp. 1593–1598, September 1998.

[14] S. He, E. Prempain, and Q. H. Wu, "An improved particle swarm optimizer for mechanical design optimization problems," *Engineering Optimization*, vol. 36, no. 5, pp. 585–605, 2004.

[15] F. S. Lobato, J. A. Sousa, C. E. Hori, and V. Steffen Jr., "Improved bees colony algorithm applied to chemical engineering system design," *International Review of Chemical Engineering*, vol. 2, no. 6, pp. 714–719, 2010.

[16] F. S. Lobato, D. L. Souza, and R. Gedraite, "A comparative study using bio-inspired optimization methods applied to controllers tuning," in *Frontiers in Advanced Control Systems*, G. Luiz de Oliveira Serra, Ed., 2012.

[17] J. Parrish, S. Viscido, and D. Grünbaum, "Self-organized fish schools: an examination of emergent properties," *Biological Bulletin*, vol. 202, no. 3, pp. 296–305, 2002.

[18] D. T. Pham, E. Kog, A. Ghanbarzadeh, S. Otri, S. Rahim, and M. Zaidi, "The bees algorithm—a novel tool for complex optimisation problems," in *Proceedings of 2nd International Virtual Conference on Intelligent Production Machines and Systems*, Elsevier, Oxford, UK, 2006.

[19] X. S. Yang, *Nature-Inspired Metaheuristic Algorithms*, Luniver Press, Cambridge, Mass, USA, 2008.

[20] X. L. Li, Z. J. Shao, and J. X. Qian, "Optimizing method based on autonomous animats: fish-swarm algorithm," *System Engineering Theory and Practice*, vol. 22, no. 11, pp. 32–38, 2002.

[21] P. Lucic and D. Teodorovic, "Bee system: modeling combinatorial optimization transportation engineering problems by swarm intelligence," in *Proceedings of TRISTAN IV Triennial Symposium on Transportation Analysis*, pp. 441–445, 2011.

[22] S. J. Zhu, Y. F. Zheng, and Y. M. Fu, "Analysis of non-linear dynamics of a two-degree-of-freedom vibration system with non-linear damping and non-linear spring," *Journal of Sound and Vibration*, vol. 271, no. 1-2, pp. 15–24, 2004.

[23] A. H. Nayfeh, *Perturbation Methods*, John Wiley & Sons, 2000.

[24] E. J. Haug, K. K. Choi, and V. Komkov, *Design Sensitivity Analysis of Structural Systems*, Academic Press, 1986.

[25] C. Andrés and S. Lozano, "A particle swarm optimization algorithm for part-machine grouping," *Robotics and Computer-Integrated Manufacturing*, vol. 22, no. 5-6, pp. 468–474, 2006.

[26] Q. He, L. Wang, and B. Liu, "Parameter estimation for chaotic systems by particle swarm optimization," *Chaos, Solitons & Fractals*, vol. 34, no. 2, pp. 654–661, 2007.

[27] K. von Frisch, *Bees: Their Vision, Chemical Senses and Language*, Cornell University Press, Ithaca, NY, USA, 1976.

[28] S. Camazine, J. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz, and E. Bonabeau, *Self-Organization in Biological Systems*, Princeton University Press, 2003.

[29] X. S. Yang, "Engineering optimizations via nature-inspired virtual bee algorithms," in *Proceedings of the 1st International Work-Conference on the Interplay Between Natural and Artificial Computation (IWINAC '05)*, J. M. Yang and J. R. Alvarez, Eds., vol. 3562 of *Lecture Notes in Computer Science*, pp. 317–323, Springer, June 2005.

[30] D. Teodorovic and M. Dell'Orco, "Bee colony optimization—a cooperative learning approach to complex transportation problems," in *Proceedings of the 10th EWGT Meeting and 16th Mini-EURO Conference*, 2005.

[31] H. S. Chang, "Converging marriage in honey-bees optimization and application to stochastic dynamic programming," *Journal of Global Optimization*, vol. 35, no. 3, pp. 423–441, 2006.

[32] A. Afshar, O. B. Haddad, M. A. Mariño, and B. J. Adams, "Honey-bee mating optimization (HBMO) algorithm for optimal reservoir operation," *Journal of the Franklin Institute*, vol. 344, no. 5, pp. 452–462, 2007.

[33] M. F. Azeem and A. M. Saad, "Modified queen bee evolution based genetic algorithm for tuning of scaling factors of fuzzy knowledge base controller," in *Proceedings of the 1st India Annual Conference (IEEE INDICON '04)*, pp. 299–303, December 2004.

[34] S. Lukasik and S. Zak, "Firefly algorithm for continuous constrained optimization task," in *Proceedings of the 1st International Conference (ICCCI '09)*, N. T. Ngugen, R. Kowalczyk, and S. M. Chen, Eds., vol. 5796 of *Lecture Notes in Artificial Intelligence*, pp. 97–100, October 2009.

[35] X. S. Yang, "Firefly algorithm, lévy flights and global optimization," in *Research and Development in Intelligent Systems XXVI*, M. Bramer, R. Ellis, and M. Petridis, Eds., pp. 209–218, Springer, London, UK, 2010.

[36] X. S. Yang, "Firefly algorithms for multimodal optimization," in *Stochastic Algorithms: Foundations and Applications*, vol. 5792, pp. 169–178, 2009.

[37] A. A. Pfeifer and F. S. Lobato, "Soluiton of singular optimal control problems using the firefly algorithm," in *Proceedings of VI Congreso Argentino de Ingeniería Química (CAIQ '10)*, 2010.

[38] T. Apostolopoulos and A. Vlachos, "Application of the firefly algorithm for solving the economic emissions load dispatch problem," *International Journal of Combinatorics*, vol. 2011, Article ID 523806, 23 pages, 2011.

[39] S. S. Madeiro, *Modal search for swarm based on density [Ph.D. dissertation]*, Universidade de Pernambuco, 2010, (Portuguese).

[40] C. R. Wang, C. L. Zhou, and J. W. Ma, "An improved artificial fish-swarm algorithm and its application in feed-forward neural networks," in *Proceedings of the 4th International Conference on Machine Learning and Cybernetics (ICMLC '05)*, pp. 2890–2894, August 2005.

[41] X. L. Li, Y. C. Xue, F. Lu, and G. H. Tian, "Parameter estimation method based on artificial fish school algorithm," *Journal of Shan Dong University*, vol. 34, no. 3, pp. 84–87, 2004.

[42] Y. Cai, "Artificial fish school algorithm applied in a combinatorial optimization problem," *Intelligent Systems and Applications*, vol. 1, pp. 37–43, 2010.

[43] A. M. A. C. Rocha, T. F. M. C. Martins, and E. M. G. P. Fernandes, "An augmented Lagrangian fish swarm based method for global optimization," *Journal of Computational and Applied Mathematics*, vol. 235, no. 16, pp. 4611–4620, 2011.

[44] W. Shen, X. Guo, C. Wu, and D. Wu, "Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm," *Knowledge-Based Systems*, vol. 24, no. 3, pp. 378–385, 2011.

[45] H. C. Tsai and Y. H. Lin, "Modification of the fish swarm algorithm with particle swarm optimization formulation and communication behavior," *Applied Soft Computing Journal*, vol. 11, no. 8, pp. 5367–5374, 2011.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 695172, 13 pages
http://dx.doi.org/10.1155/2013/695172

Research Article

Surrogate-Assisted Multiobjective Evolutionary Algorithms for Structural Shape and Sizing Optimisation

Tawatchai Kunakote and Sujin Bureerat

Department of Mechanical Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand

Correspondence should be addressed to Sujin Bureerat; sujbur@kku.ac.th

Received 14 February 2013; Revised 4 June 2013; Accepted 5 June 2013

Academic Editor: Guilin Yang

The work in this paper proposes the hybridisation of the well-established strength Pareto evolutionary algorithm (SPEA2) and some commonly used surrogate models. The surrogate models are introduced to an evolutionary optimisation process to enhance the performance of the optimiser when solving design problems with expensive function evaluation. Several surrogate models including quadratic function, radial basis function, neural network, and Kriging models are employed in combination with SPEA2 using real codes. The various hybrid optimisation strategies are implemented on eight simultaneous shape and sizing design problems of structures taking into account of structural weight, lateral bucking, natural frequency, and stress. Structural analysis is carried out by using a finite element procedure. The optimum results obtained are compared and discussed. The performance assessment is based on the hypervolume indicator. The performance of the surrogate models for estimating design constraints is investigated. It has been found that, by using a quadratic function surrogate model, the optimiser searching performance is greatly improved.

1. Introduction

Since they have been invented for several decades, the implementation of evolutionary optimisers on a wide variety of real-world design problems has successfully been made in numerous fields. The algorithms are attractive and popular as they can responsd to unavoidable disadvantages of classical mathematical programming. The evolutionary algorithms (EAs) can deal with almost all kinds of design problem as their search mechanisms, to some extent, rely on randomisation and need no function derivatives, for example, [1–4]. Recently, single-objective evolutionary methods that have outstanding performance in several applications are real-code ant colony optimisation (ACOR) [5], covariance matrix adaptation evolution strategy (CMA-ES) [6], and differential evolution (DE) [7]. The methods are robust and capable of reaching global optima. Most importantly, for multiple-objective design cases, they can be used to explore a Pareto optimal front of the problem within one simulation run. However, EAs have some unacceptable drawbacks that are a complete lack of consistency and low convergence rate. As a result, the obtained design results are considered near local optima for single objective cases and approximate Pareto fronts for multiobjective design [8].

With those aforementioned shortcomings, it is difficult or even impossible to employ the methods for solving the design problems with expensive function evaluation or limited number of function values available. Such obstructions can however be alleviated by introducing a surrogate model (SM) to an evolutionary optimisation procedure. With the use of such a model, an approximate function can be constructed, and inexpensive function prediction can be achieved. The hybridisation of EAs and SM for design optimisation has been studied since the last decade, and it has been found to greatly enhance the performance of EAs. Many researchers and engineers have focused their efforts on this research issue. The design problems can have single or multiple design objectives, for example, in [9–14] while mostly the demonstration problems are unconstrained or box-constrained. The hybrid EAs with SMs have been implemented on a wide variety of engineering application, for example, aerodynamics [15–17], heat and mass transfer [18–20], and structures [13, 14, 21–23]. Design problems requiring a surrogate-assisted optimisation may be roughly categorised to have 9 classes as given in [10]. Surrogate-assisted EAs not only employ function approximation but also use some mathematical programming to enhance the searching performance [24, 25]. A comprehensive survey on the surrogate models used for enhancing EAs performance can be found in [26–29].

The surrogate models commonly employed include polynomial regression techniques [11, 20, 22, 26, 27, 30], radial basis function interpolation (RBF) [9, 11, 12, 16, 18, 27], artificial neural network (ANN) [9, 15, 23, 26], support vector machines [31], moving least square method [32], and Gaussian process model (Kriging model) [10, 16, 25–27]. Recent alternative surrogate-based approaches are concerned with the approximation of field vectors (such as velocity and pressure fields for computational fluid dynamics and displacement and stress fields for structural analysis) rather than approximating objective and constraint functions directly. The proper orthogonal decomposition (POD) technique is employed for this task [33–35]. With this concept, a designer can deal with more complicated design problems such as design optimisation when aeroelasticity is taken into consideration. Apart from that, there exist optimisers with the use of grid-based computing integrating nongeneration evolutionary algorithms, surrogate models, local search, and Lamarckian learning process together. This approach is termed asynchronous metamodel-based evolutionary algorithms [36, 37]. A technique used in the design of computer experiment phase is usually a Latin hypercube sampling technique. Recently, there have been several papers focusing on the development of optimum Latin hypercube sampling [38, 39]. Another interesting approach is an infill sampling technique [40].

The most challenging task for surrogate-based optimisation is to improve the convergence rate. To the authors' best knowledge, the work in the literature mostly demonstrated applying the design approaches to unconstrained or box-constrained optimisation which is simple to handle by using EAs. Nevertheless, there are usually many nonlinear constraints in the real-world design problems particularly in structural optimisation. The approximation of objective function, although being inaccurate, can lead to real optima (referred to as "Blessing of Uncertainty" in [11]). On the other hand, in estimating constraint functions, the approximate function needs to be as precise as possible so that the evolutionary search can land in the real feasible region. Inaccurate constraint function approximation can lead the optimiser to an undesirable design space.

In this paper, the hybridisation of the well-established multiobjective evolutionary algorithm (MOEA), namely, strength Pareto evolutionary algorithm [41] and some popular surrogate models is developed. The work is concerned with the studies of using several SMs for performance enhancement of multiobjective evolutionary algorithms in solving structural constrained optimisation. A number of surrogate models including quadratic function, radial basis function, neural network, and Kriging models are employed in combination with SPEA2 using real codes. The various hybrid optimisation strategies are implemented on eight simultaneous shape and sizing design problems of a structure having design functions as structural weight, lateral bucking, natural frequency, and stress. Structural analysis is carried out by using finite element analysis (FEA). The optimum results obtained are compared and discussed. The performance of SMs for estimating design constraints is investigated. It has been found that, by using a quadratic surrogate model, the searching performance of SPEA2 is greatly improved.

2. Hybridisation of SPEA2 and a Surrogate Model

There have been several ways to properly integrate SMs into EAs. The hybrid algorithm proposed in this paper is similar to what is presented in [18]. The SPEA2 with real codes is employed as the main MOEA procedure. The surrogate-assisted SPEA is illustrated in Figure 1. The computational steps in the figure can be separated into two parts as the main SPEA procedure and the SPEA subproblem that uses an approximation model. The main SPEA is slightly modified from the original version presented in [41]. The real code recombination and mutation are similar to the work presented in [18]. The procedure starts with an initial population, an (empty) Pareto archive, and some other predefined parameters, for example, an external archive size and recombination and mutation rates. The fitness of the individuals in the population is then assigned by performing actual function evaluation. The current population and their corresponding fitness values are brought to the surrogate environment as some of them are picked to build a surrogate model. Having such an approximation model, the SPEA is employed to solve the optimisation problem with approximate function evaluation.

The nondominated solutions from this stage are put back to the main SPEA procedure where their actual fitness values are determined. The current population, the external archive, and the nondominated solutions from performing a surrogate-based optimisation are put together while their nondominated solutions are determined. The external archive is then updated by replacing it with these nondominated solutions. In case that the number of the nondominated solutions exceeds the predefined archive size, some of them will be removed from the archive by means of the nearest neighbourhood method [41]. Subsequently, the so-called binary tournament selection is performed to select some members in the newly updated archive to reproduce a set of offspring. The algorithm goes back to the step of fitness function evaluation and is repeated until the termination criterion is fulfilled.

Design solutions are selected from a current population to build a surrogate model based upon the idea that they should be evenly distributed throughout the design space. The proposed approach is operated on the domain of objective functions, which is somewhat similar to the adaptive grid algorithm, a Pareto archiving technique used in the Pareto archive evolution strategy [8]. The procedure to select the solutions for surrogate modelling is illustrated in Figure 2. On the objective domain, all of the design points in the current population are plotted, and a proper grid is generated covering all the points. A solution that is the closest to the centre of each mesh is selected for function estimation.

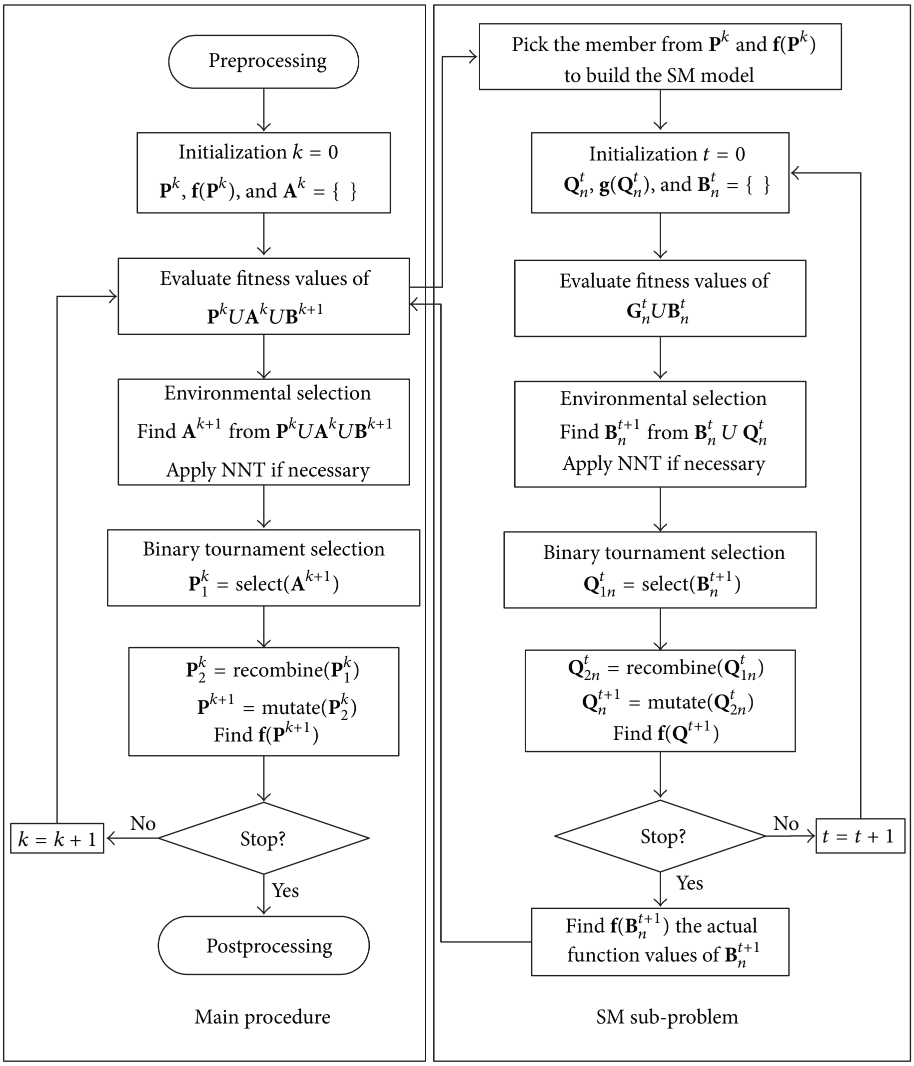


Figure 1: Flowchart for surrogate-assisted SPEA.

3. Surrogate Models

Let $y = f(\mathbf{x})$ are a function of a design vector $\mathbf{x}$ sized $n \times 1$. Given a set of design solutions (or sampling points) $\mathbf{X} = [\mathbf{x}^1, \ldots, \mathbf{x}^N]$ and their corresponding function values $\mathbf{Y} = [y_1, \ldots, y_N]$ selected during an evolutionary optimisation process, a surrogate or approximation model is constructed by means of curve fitting or interpolation. Approximation models used in this study are as follows.

3.1. Polynomial Regression. The most commonly used polynomial surrogate model or a response surface model (RSM) is of the second-order polynomial or quadratic model, which can be expressed as

$$\overline{y} = \beta_0 + \sum \beta_i x_i + \sum \beta_i x_i x_j, \tag{1}$$

where β_i for $i = 0, \ldots, (n+1)(n+2)/2$ are the polynomial coefficients to be determined. The coefficients can be found by using a regression or least square technique [26].

3.2. Kriging Model. A Kriging model (also known as a Gaussian process model) used in this paper is the famous MATLAB toolbox named design and analysis of computer experiments (DACE) [42]. The estimation of function can be thought of as the combination of global and local approximation models, that is,

$$y(\mathbf{x}) = \overline{f}(\mathbf{x}) + Z(\mathbf{x}), \tag{2}$$

where $\overline{f}(\mathbf{x})$ is a global regression model while $Z(\mathbf{x})$ is a stochastic Gaussian process with zero mean and nonzero covariance representing a localised deviation. In this work, a linear function is used for a global model, which can be expressed as

$$\overline{f} = \beta_0 + \sum_{i=1}^{n} \beta_i x_i = \boldsymbol{\beta}^T \mathbf{f}, \tag{3}$$

where $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_n]^T$, $\mathbf{f} = \mathbf{f}(\mathbf{x}) = [1, x_1, x_2, \ldots, x_n]^T$. The covariance of $Z(\mathbf{x})$ is expressed as

$$\text{Cov}(Z(\mathbf{x}^p), Z(\mathbf{x}^q)) = \sigma^2 \mathbf{R}[R(\mathbf{x}^p, \mathbf{x}^q)] \tag{4}$$

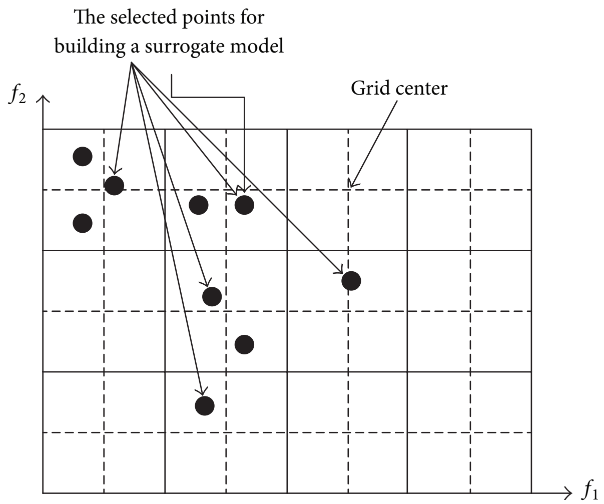


FIGURE 2: Selection of sampling points for constructing a surrogate model.

for $p, q = 1, \ldots, N$ where R is the correlation function between any two of the N design points, and $\mathbf{R}$ is the symmetric correlation matrix size $N \times N$ with the unity diagonal [26]. The correlation function used herein is

$$R\left(\mathbf{x}^p, \mathbf{x}^q\right) = \exp\left(-\left(\mathbf{x}^p - \mathbf{x}^q\right)^T \boldsymbol{\theta}\left(\mathbf{x}^p - \mathbf{x}^q\right)\right), \quad (5)$$

where θ_i are the unknown correlation parameters to be determined by means of the maximum likelihood method. Having found $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, the Kriging predictor can be achieved as

$$\overline{y} = \mathbf{f}\left(\mathbf{x}\right)^T \boldsymbol{\beta} + \mathbf{r}^T\left(\mathbf{x}\right)\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{F}\boldsymbol{\beta}\right), \quad (6)$$

where $\mathbf{F} = [\mathbf{f}(\mathbf{x}^1), \mathbf{f}(\mathbf{x}^2), \ldots, \mathbf{f}(\mathbf{x}^n)]^T$ and $\mathbf{r}^T(\mathbf{x}) = [R(\mathbf{x}, \mathbf{x}^1), R(\mathbf{x}, \mathbf{x}^2), \ldots, R(\mathbf{x}, \mathbf{x}^N)]$. For more details, see [42].

3.3. Radial Basis Function Interpolation. The radial basis function interpolation has been used in a wide range of applications such as integration between aerodynamic and finite element grids in aeroelastic analysis [43]. The use of such a model for surrogate-assisted evolutionary optimisation is said to be commonplace [9, 18, 27]. The approximate function can be written as

$$\overline{y}\left(\mathbf{x}\right) = \sum_{i=1}^{N} \alpha_i K\left(\left\|\mathbf{x} - \mathbf{x}^i\right\|\right), \quad (7)$$

where α_i are the coefficients to be determined, and K is a radial basis kernel. (Here it is set to be linear splines.) The coefficients can be found from the N sampling points as detailed in Srisomporn and Bureerat 2008 [18].

3.4. Neural Network. Artificial neural network (ANN) is developed in response to the need to use complicated function expression rather than a simple quadratic or radial basis function. The model consists of interconnecting neurons or nodes that somewhat mimic the constructs of biological neurons as shown in Figure 3. ANN has been implemented on a wide variety of real world applications. The use of ANN for surrogate-assisted optimisation can be found in the literature [9]. The neural network model used in this paper is the feedforward multilayer perceptrons. The details of the network in this work are the learning algorithm = *backpropagation* and network training function = *Levenberg-Marquardt backpropagation.*

TABLE 1: Design variables and bounds for F1–F6.

Design variables	Bounds (mm)
$x_1 = l_d$	$0 \le l_d \le 60$
$x_2 = l_p$	$100 \le l_p \le 200$
$x_3 = r_{p1}$	$15 \le r_{p1} \le 35$
$x_4 = r_{p2}$	$15 \le r_{p2} \le 30$
$x_5 = \text{th}$	$3 \le \text{th} \le 4.5$

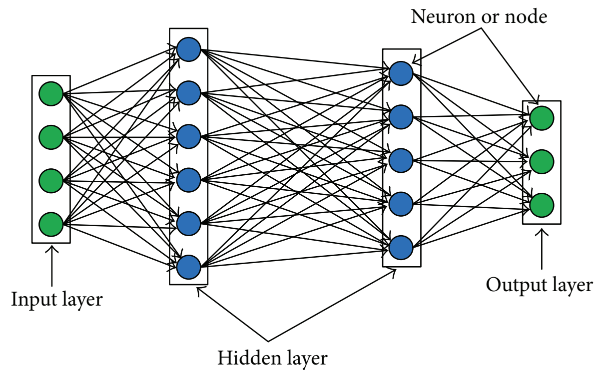


FIGURE 3: Neural network.

In this work, a simple heuristic algorithm is used to construct a network topology. The procedure starts with a network with two hidden layers while the number of nodes N_R on each layer is randomly chosen with the bound $N_R \in [15, 50]$. The transfer function for the hidden layer can be either *hyperbolic tangent sigmoid* or *logarithmic sigmoid* based on randomisation. If the termination criterion (training goal) is not met, randomly increase a number of nodes in the layers. If the number of nodes reaches the upper bound $N_R = 50$, add the third hidden layer to the network where a transfer function and node number are randomly chosen. In case that the stopping criterion is not fulfilled, increase the number of nodes. The algorithm repeats until the stopping criterion is met.

4. Testing Problems

Six testing multiobjective design problems are posed to design a torque arm [44] structure shown in Figure 4. Since it is a plate-like structure under inplane loading, design criteria to be taken into consideration are weight, stress, buckling, and natural frequency. The function calculation can be carried out by using FEA where shell elements are employed. The design variables and their bound constraints are given in Table 1 where other dimensions are set as $l_d = 30$ mm, $l_p = 190$ mm, $r_{p1} = 26$ mm, and $r_{p2} = 27$ mm.

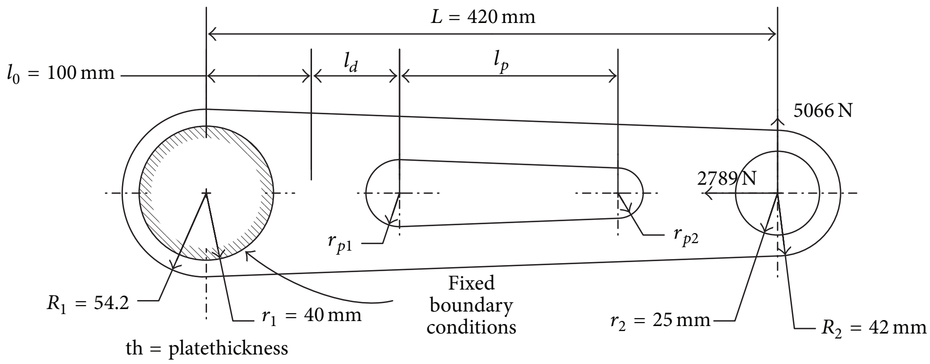


Figure 4: Torque arm.

The six biobjective optimisation problems are termed F1, F2, F3, F4, F5, and F6, respectively, and can be expressed as:

objective function:

$$\min\ f_1(\mathbf{x}),\ \max f_2(\mathbf{x})$$
$$\text{F1}: f_1(\mathbf{x}) = w, \quad f_2(\mathbf{x}) = \omega$$
$$\text{F2}: f_1(\mathbf{x}) = w, \quad f_2(\mathbf{x}) = \lambda$$
$$\text{F3}: f_1(\mathbf{x}) = w, \quad f_1(\mathbf{x}) = \lambda + \frac{\omega}{\omega_0}$$

subject to

$$\sigma_{\text{von}} - \frac{\sigma_y}{S_F} \leq 0$$
$$1 - \lambda \leq 0,$$

objective function:

$$\min f_1(\mathbf{x}),\quad \max f_2(\mathbf{x})$$
$$\text{F4}: f_1(\mathbf{x}) = w, \quad f_2(\mathbf{x}) = \omega$$
$$\text{F5}: f_1(\mathbf{x}) = w, \quad f_2(\mathbf{x}) = \lambda$$
$$\text{F6}: f_1(\mathbf{x}) = w, \quad f_2(\mathbf{x}) = \lambda + \frac{\omega}{\omega_0}$$

subject to

$$\sigma_{\text{von}} - \frac{\sigma_y}{S_F} \leq 0$$
$$1 - \lambda \leq 0$$
$$\omega - \omega_0 \leq 0, \tag{8}$$

where w = structural weight, ω = first mode natural frequency, λ = lateral bucking factor (it is the ratio of critical load to applied load herein), σ_{von} = maximum von Mises stress on the structure, and $\omega_0 = 7.2688$ rad/s. Furthermore, material properties are Young's modulus $E = 60 \times 10^6$ N/cm^2, Poisson's ratio $\nu = 0.3$, yield stress $\sigma_y = 80000$ N/cm^2, density $\rho = 0.00718$ kg/cm^3, and safety factor $S_F = 1.5$.

In fact, the F1, F2, and F3 problems are similar to the F4, F5, and F6 problems, respectively, except for the introduction of natural frequency constraints to the last three problems. This means that the performance of several surrogate models in estimating a natural frequency during an optimisation procedure can be examined. The bound constraints in Table 1 are not presented in the design problems as they can be dealt with in the SPEA procedure. It should be noted that the finite element analysis for these design studies is not time consuming as the work focuses on a comparative study of the several multiobjective optimisers.

Since the F1–F6 problems are small-scale, other two test problems are posed so as to investigate the capability of those surrogate models when dealing with large-scale problems. Two design problems of a simple wing-box displayed in Figure 5 are expressed as

objective function:

$$\min f_1(\mathbf{x}),\ \max f_2(\mathbf{x})$$
$$\text{F7}: f_1(\mathbf{x}) = w, \quad f_2(\mathbf{x}) = -\lambda$$
$$\text{F8}: f_1(\mathbf{x}) = w, \quad f_2(\mathbf{x}) = -(\omega_1 + \cdots + \omega_5) \tag{9}$$

subject to

$$\sigma_{\text{von}} - \sigma_y \leq 0$$
$$1 - \lambda \leq 0,$$

where ω_i is the ith mode natural frequency of the wing-box. The wing-box is made up of Aluminium with Young modulus $E = 70$ GPa, Poisson's ratio $\nu = 0.34$, density $\rho = 2700$ kg/m^3, and yield strength $\sigma_{yt} = 100$ MPa. The structure consists of 2 spars, 4 ribs, and 4 skins (front, rear, top, and bottom skins) resulting in 52 wing segments as shown. The thickness of each segment is assigned to be a design variable; therefore, there are 52 design variables with the lower and upper bounds being 0.002 m and 0.010 m respectively. This can be classified as a large-scale sizing optimisation problem. The wing is subject to static loads on the front spar as shown in Figure 5. Shell elements are used for finite element analysis similarly to F1–F6.

The multiobjective optimisers used in this paper are named as follows:

(i) SPEA-O the original SPEA2 without using a surrogate model,

(ii) SPEA-Q SPEA using a quadratic response surface model,

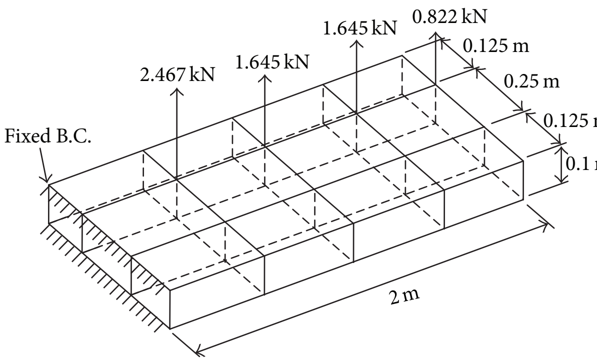


Figure 5: Simple wing-box.

(iii) SPEA-R SPEA using the radial basis function interpolation,

(iv) SPEA-N SPEA using a neural network model,

(v) SPEA-K SPEA using a Kriging model.

Each evolutionary algorithm is employed to solve each problem ten runs with 30 generations (50 for F7-F8) and a population size of 20 (50 for F7-F8) for SPEA-O. For the algorithms using an approximation model, the number of generations and population size used in the main procedure is 15 (25 for F7-F8) and 20 (50 for F7-F8), respectively, while on each generation additional 20 (50 solutions for F7-F8) solutions are taken from optimising a surrogate subproblem. In the surrogate subproblem, the number of generations and population size is set to be 100 and 100, respectively. The external Pareto archive size is set to be 100. During the optimisation process, constraints are handled by using the approach presented in [45]. It should be noted that there have been a number of techniques used to handle design constraints in evolutionary optimisation [46]. We choose the aforementioned technique because it is simple and sufficiently effective.

Table 2: Comparison of normalised average hypervolume.

	SPEA-O	SPEA-Q	SPEA-R	SPEA-N	SPEA-K
F1	0.429	0.704	1.000	0.000	0.523
F2	0.000	1.000	0.052	0.917	0.692
F3	0.577	0.995	0.745	0.000	1.000
F4	0.457	0.916	1.000	0.000	0.497
F5	0.000	0.976	0.640	0.161	1.000
F6	0.000	0.673	0.243	0.007	1.000
$\sum$ F1–F6	1.463	5.264	3.679	1.084	4.713
F7	0.811	1.000	0.724	N/A	0.000
F8	0.337	1.000	0.453	N/A	0.000
$\sum$ F7-F8	1.148	2.000	1.177	N/A	0.000

Table 3: Performance matrix of F1.

	SPEA-O	SPEA-Q	SPEA-R	SPEA-N	SPEA-K
SPEA-O	0	0	1	0	0
SPEA-Q	0	0	1	0	0
SPEA-R	0	0	0	0	0
SPEA-N	1	1	1	0	1
SPEA-K	0	0	1	0	0
Total	1	1	4	0	1
Ranking	2	2	1	5	2

Table 4: Comparison of ranking scores by t-test.

	SPEA-O	SPEA-Q	SPEA-R	SPEA-N	SPEA-K
F1	2	2	1	5	2
F2	4	1	4	1	1
F3	1	1	1	5	1
F4	3	1	1	5	3
F5	4	1	3	4	1
F6	3	2	3	3	1
$\sum$ F1–F6	17	8	13	23	9
F7	1	1	1	N/A	4
F8	2	1	2	N/A	4
$\sum$ F7-F8	3	1	3	N/A	8

5. Results and Discussion

The performance comparison of the five MOEAs for solving the eight design problems is based on the hypervolume indicator. Two types of performance assessment are conducted as detailed in Tables 2–4. In Table 2, each cell is the normalised mean value of 10 nondominated front hypervolumes obtained from using the optimiser on the corresponding column for solving the problem on the corresponding row. For each design problem, the best and worst mean values are normalised to be "1" and "0", respectively. The total normalised mean values given in the table are used for evaluating the overall performance of the surrogate-assisted SPEAs. From the results, the best algorithm for F1 is SPEA-R while the second best is SPEA-Q. SPEA-Q is the best for F2 while the second best is SPEA-N. The best optimisers for F3, F4, F5, and F6 are SPEA-K, SPEA-R, SPEA-K, and SPEA-K, respectively, whereas the second best for those 4 design problems is SPEA-Q. SPEA-O without using a surrogate model is the worst performer for F2, F5, and F6 while SPEA-N is the worst performer for F1, F3, and F4. The overall best performer based on the evaluation is SPEA-Q whereas the second best is SPEA-K. For the large-scale problems, only SPEA-Q can surpass SPEA-O. SPEA-R is said to be equal to SPEA-O while SPEA-K gives that worst performance. Note that SPEA-N cannot be applied to these problems as it takes excessively long running time for network training. It can be said that only the quadratic response surface model is useful for the large-scale wing-box design.

The second performance assessment is based on the statistical t-test. For each design problem, to compare the 5 MOEAs, a performance matrix $\mathbf{T}$ sized 5×5 whose elements are full of "0" is generated. For the element T_{ij} of the matrix, if the mean value of the 10 hypervolumes obtained from method j is significantly larger than that obtained from method i based on the statistical t-test at 95% confidence level, the value of T_{ij} is set to be one. Table 3 shows the performance comparison of the problem F1. Having had the performance matrix, the 6th row on the table displays the total scores of each optimiser. The ranking is made in such

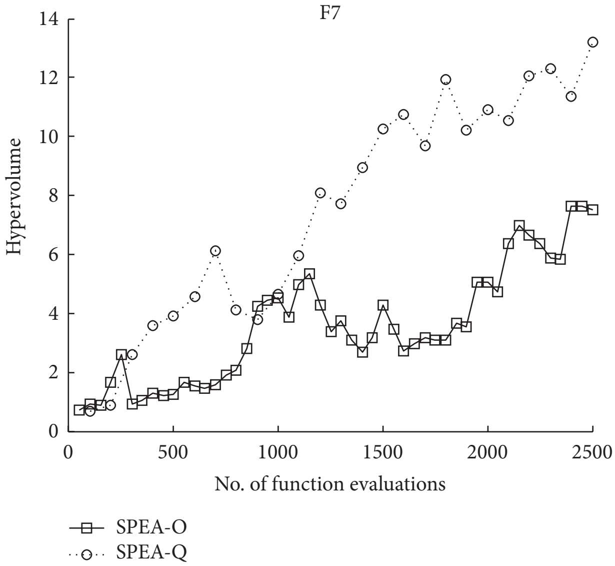


Figure 6: SPEA-O versus SPEA-Q for F7.

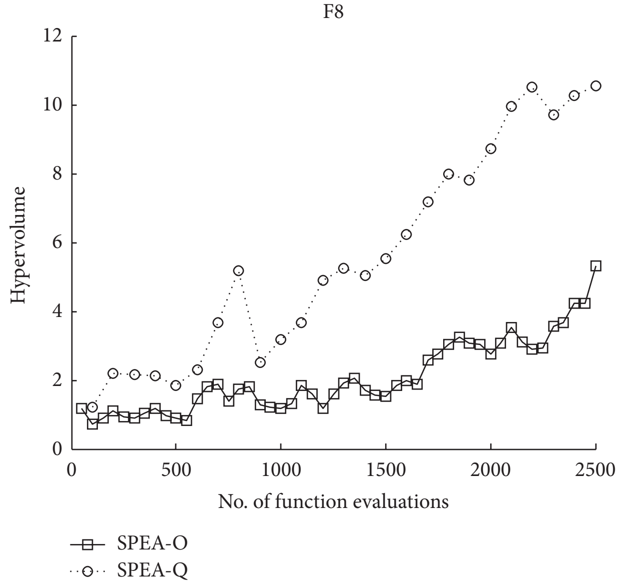


Figure 7: SPEA-O versus SPEA-Q for F8.

a way that the best method is ranked as 1 while the worst is ranked as 5. The overall performance assessment based on the t-test is given in Table 4 where the last row on the table shows the total score of all the MOEAs. The lower the score the better the performer. Similar to the first assessment, the overall best method for F1–F6 based on this assessment is SPEA-Q while the close second best is SPEA-K. The worst method is SPEA-O without using a surrogate model. For the large-scale test problems, the only surrogateassisted method that is superior to the non-surrogate SPEA is SPEA-Q.

Figures 6 and 7 show the search history as plots of hypervolume against the number of function evaluations of SPEA-O and SPEA-Q for solving F7 and F8, respectively. The values of front hypervolume can fluctuate due to the use of a Pareto archiving technique to remove some nondominated solutions from the Pareto archive. From the figures, it can be seen that SPEA-Q requires approximately half of the total number of function evaluations employed by SPEA-O to have equally good nondominated fronts.

The best approximate Pareto front of F1 obtained after numerous optimisation runs of the surrogate-assisted SPEAs is chosen and plotted in Figure 8 whereas the structures of some selected design points are shown in Figure 9. Figure 10 displays the best front of the F2 problem where some selected design solutions are illustrated in Figure 11. The best front and some selected optimal structures of the design problem F3 are displayed in Figures 12 and 13, respectively. Similarly, the best approximate fronts and some selected design solutions of F4, F5, and F6 are illustrated in Figures 14, 15, 16, 17, 18, and 19. Figures 20 and 21 show the best fronts obtained from using the various SPEAs for solving F7 and F8, respectively. It can be seen that the best fronts are from SPEA-Q, and SPEA-O. The SPEA-Q front does not totally dominate that of SPEA-O but they overlap each other while the SPEA-Q front has obviously, greater front extension.

According to the aforementioned performance assessments, it can be said that SPEA-R is the best algorithm in

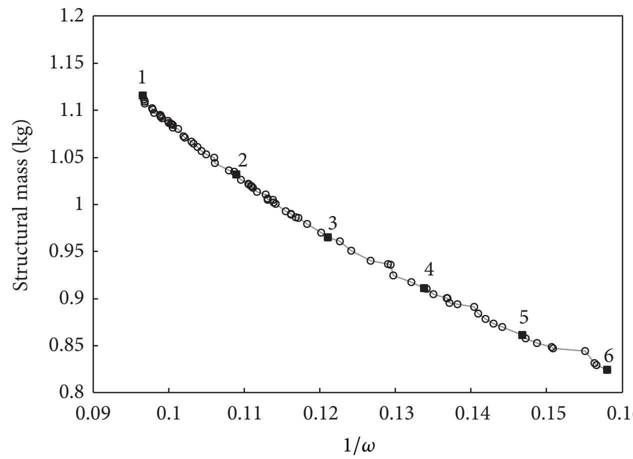


Figure 8: Approximate Pareto front of F1.

the cases of the F1 design problem while the close second best is SPEA-Q. When adding a natural frequency constraint to F1 becoming the F4 problem, the overall best and second best methods are still SPEA-R and SPEA-Q, respectively. For the F2 case and the F5 design problem which is F2 with additional natural frequency constraint, the top two performers are SPEA-Q and SPEA-K. The same can be said for the case of F3 while SPEA-K outperforms the others for the F6 design problem. The surrogate models, that is, the quadratic response surface model, the radial basis function interpolation, and the Kriging model, are said to enhance the searching performance of SPEA for the small-scale problems. The addition of constraints to the design problems can cause the searching performance of the hybrid algorithms. The SPEA using ANN is inferior to the other surrogate-assisted methods or even the original SPEA2 without using a surrogate model because it requires efficient network topology optimisation before being used to estimate functions while in this paper the simple heuristic approach is employed.

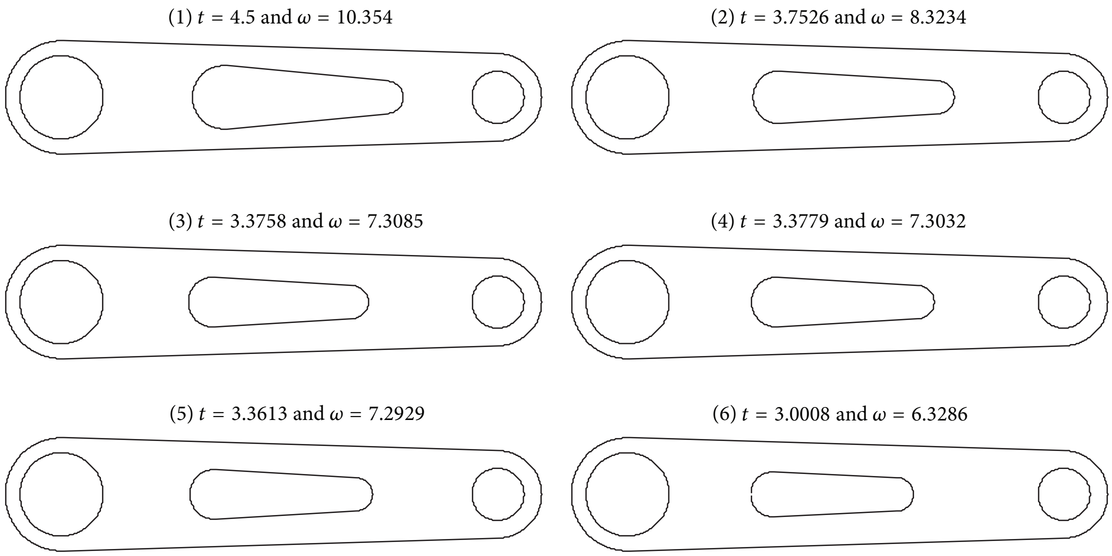


FIGURE 9: Selected design points from Figure 8.

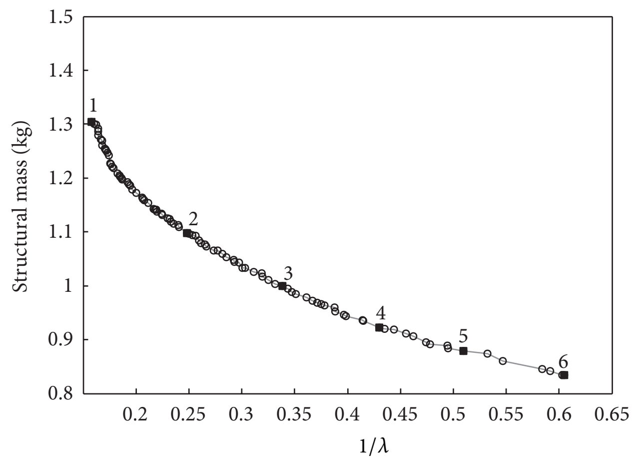


FIGURE 10: Approximate Pareto front of F2.

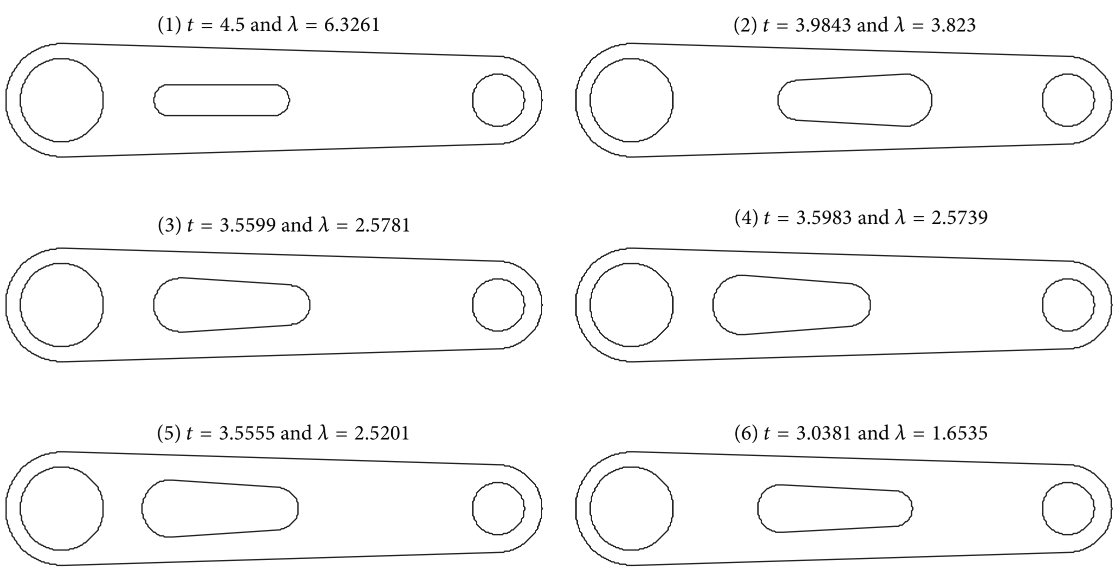


FIGURE 11: Selected design solutions from Figure 10.

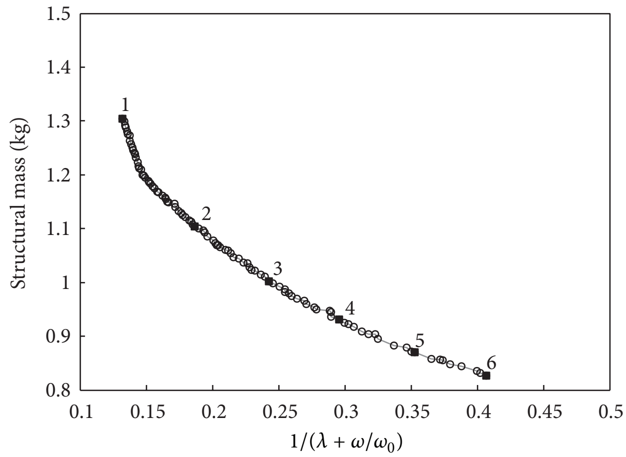


Figure 12: Approximate Pareto front of F3.

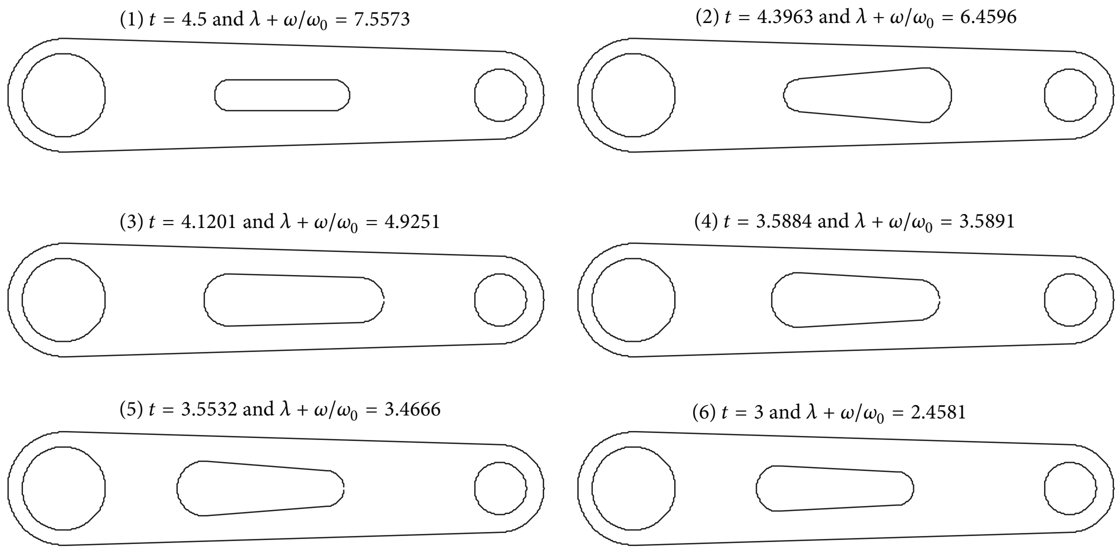


Figure 13: Selected design solutions from Figure 12.

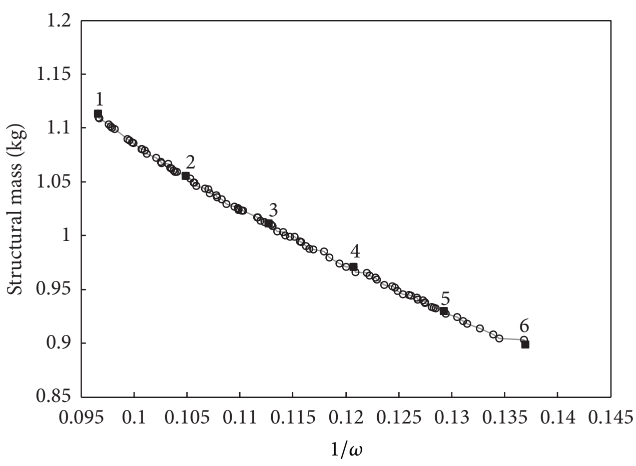


Figure 14: Approximate Pareto front of F4.

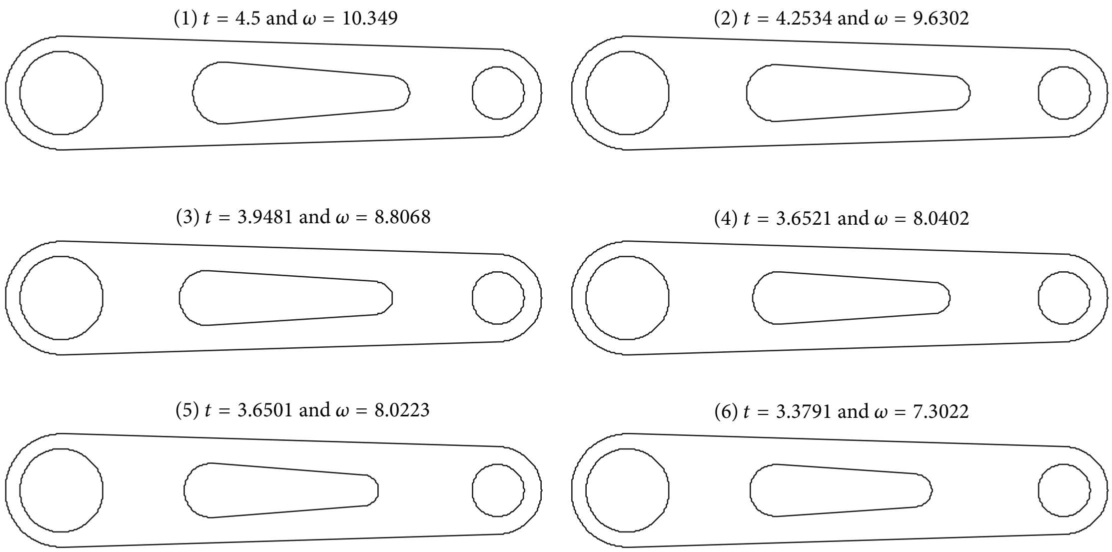


FIGURE 15: Selected design solutions from Figure 14.

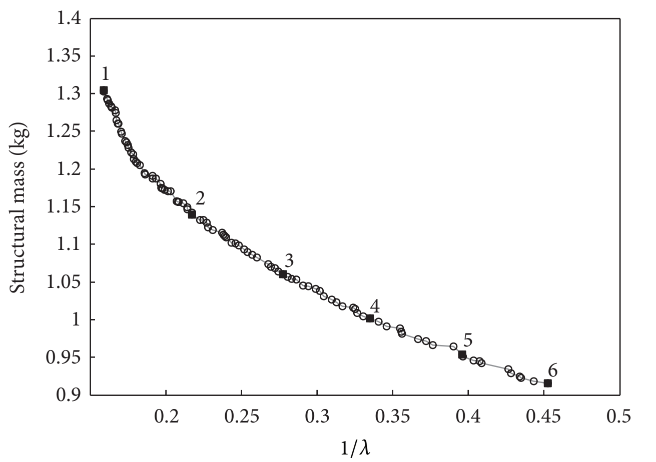


FIGURE 16: Approximate Pareto front of F5.

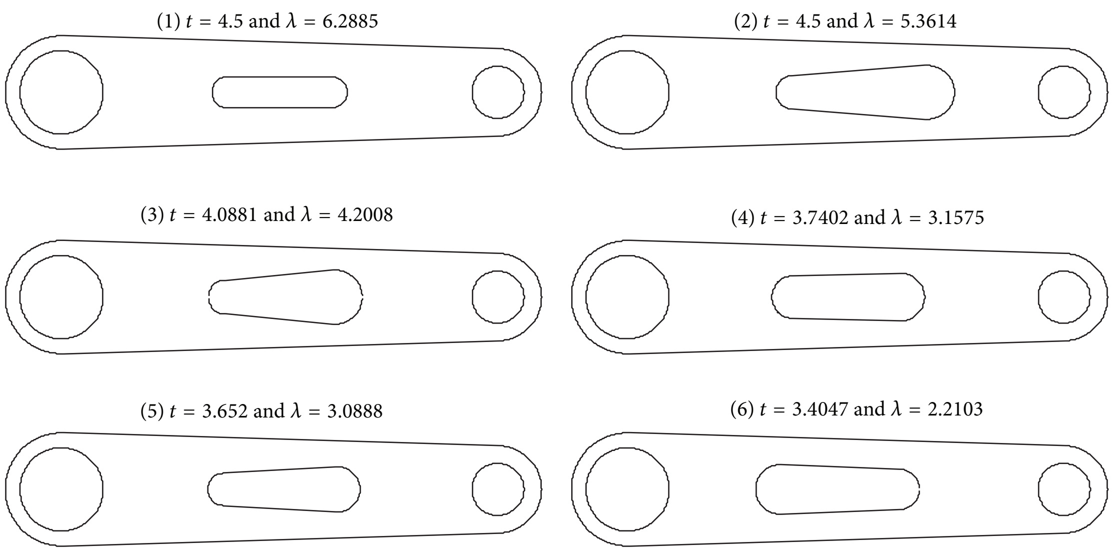


FIGURE 17: Selected design solutions from Figure 16.

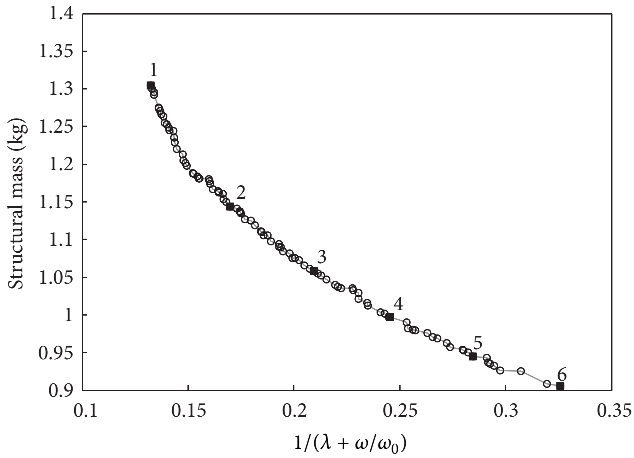


Figure 18: Approximate Pareto front of F6.

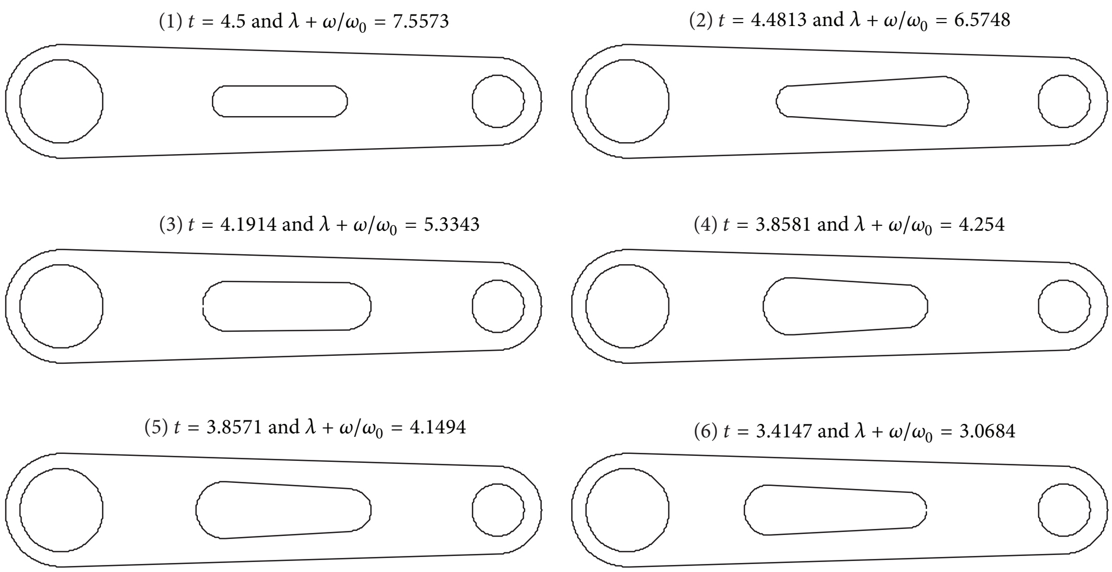


Figure 19: Selected design solutions from Figure 18.

More investigation on this issue would greatly improve the performance of SPEA-N. In cases of large-scale problems, only SPEA using a quadratic function is superior to the nonsurrogate SPEA.

6. Conclusions

The use of multiobjective evolutionary algorithm for the design of a torque arm leads to multiple optimum structures for decision making. The hybridisation of SPEA and several surrogate models is developed. The various surrogate models including the quadratic regression technique, the radial basis function interpolation, and the Kriging model can help improving the searching performance of the strength Pareto evolutionary algorithm. For a design problem having mass and natural frequency as objective functions and stress and buckling as constraints, SPEA using RBF is the best method. The SPEA using a quadratic regression is the best for the design case of F2. For the other multiobjective design problems, the best and the second best performers are SPEA using the Kriging model and the quadratic RSM, respectively. Nevertheless, when considering all of the design problems the overall top performer is SPEA2 using the quadratic regression while the close second best uses the Kriging model. Moreover, SPEA using a quadratic response surface model is the only strategy that is useful for solving large-scale problems. The use of ANN for approximating constrained optimisation problems is not effective as a proper network topology optimisation needs to be constructed before entering the surrogate subproblem. It is however possible to enhance the performance of the ANN-based optimiser if an efficient network topology optimisation is performed during the optimisation process.

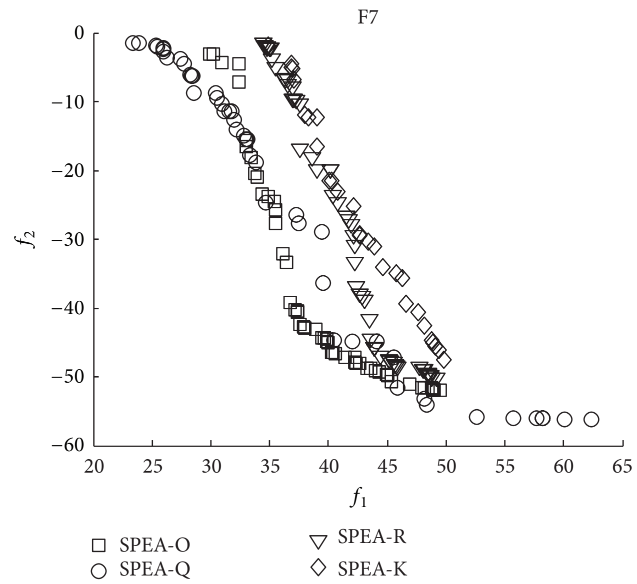


Figure 20: Approximate Pareto front of F7.

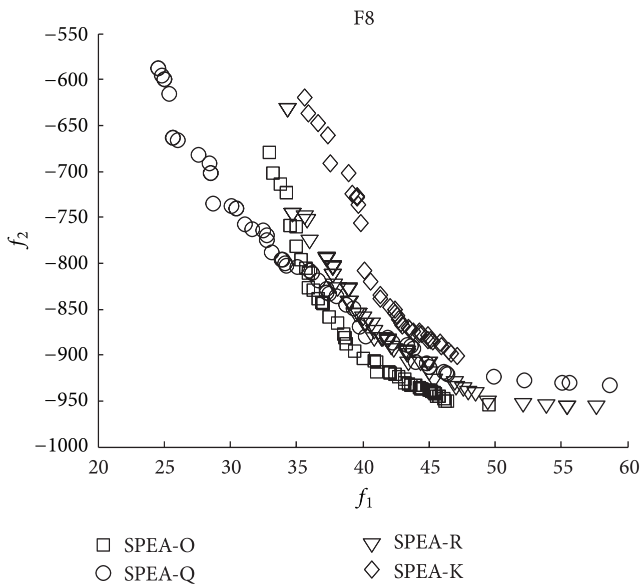


Figure 21: Approximate Pareto front of F8.

Acknowledgment

The authors are graceful for the support from the Thailand Research Fund (RSA5280006). The first author also wishes to thank the office of higher education commission, Thailand, for the CHE-PHD scholarship.

References

[1] F. Zong, H. Lin, B. Yu, and X. Pan, "Daily commute time prediction based on genetic algorithm," *Mathematical Problems in Engineering*, vol. 2012, Article ID 321574, 20 pages, 2012.

[2] I. Yao and D. Han, "Improved barebones particle swarm optimization with neighborhood search and its application on ship design," *Mathematical Problems in Engineering*, vol. 2013, Article ID 175848, 12 pages, 2013.

[3] Y. Xu, P. Fan, and L. Yuan, "A simple and efficient artificial bee colony algorithm," *Mathematical Problems in Engineering*, vol. 2013, Article ID 526315, 9 pages, 2013.

[4] S. Talatahari, E. Khalili, and S. M. Alavizadeh, "Accelerated particle swarm for optimum design of frame structures," *Mathematical Problems in Engineering*, vol. 2013, Article ID 649857, 6 pages, 2013.

[5] K. Socha and M. Dorigo, "Ant colony optimization for continuous domains," *European Journal of Operational Research*, vol. 185, no. 3, pp. 1155–1173, 2008.

[6] N. Hansen, S. D. Müller, and P. Koumoutsakos, "Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)," *Evolutionary Computation*, vol. 11, no. 1, pp. 1–18, 2003.

[7] R. Storn and K. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[8] J. D. Knowles and D. W. Corne, "Approximating the nondominated front using the Pareto Archived Evolution Strategy," *Evolutionary Computation*, vol. 8, no. 2, pp. 149–172, 2000.

[9] M. Farina and P. Amato, "Linked interpolation-optimization strategies for multicriteria optimization problems," *Soft Computing*, vol. 9, pp. 54–65, 2005.

[10] J. Knowles, "ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 1, pp. 50–66, 2006.

[11] Z. Zhou, Y. S. Ong, M. H. Lim, and B. S. Lee, "Memetic algorithm using multi-surrogates for computationally expensive optimization problems," *Soft Computing*, vol. 11, no. 10, pp. 957–971, 2007.

[12] R. G. Regis and C. A. Shoemaker, "Local function approximation in evolutionary algorithms for the optimization of costly functions," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 5, pp. 490–505, 2004.

[13] N. Pholdee and S. Bureerat, "Surrogate-assisted evolutionary optimizers for multiobjective design of a torque arm structure," *Applied Mechanics and Materials*, vol. 101-102, pp. 324–328, 2011.

[14] N. Pholdee and S. Bureerat, "Passive vibration control of an automotive component using evolutionary optimisation," *Journal of Research and Applications in Mechanical Engineering*, vol. 1, pp. 19–23, 2011.

[15] K. C. Giannakoglou, D. I. Papadimitriou, and I. C. Kampolis, "Aerodynamic shape design using evolutionary algorithms and new gradient-assisted metamodels," *Computer Methods in Applied Mechanics and Engineering*, vol. 195, no. 44–47, pp. 6312–6329, 2006.

[16] Y. S. Ong, K. Y. Lum, P. B. Nair, D. M. Shi, and Z. K. Zhang, "Global convergence of unconstrained and bound constrained surrogate-assisted evolutionary search in aerodynamic shape design," in *Proceedings of the Congress on Evolutionary Computation (CEC '03)*, pp. 1856–1863, Canberra, Australia, December 2003.

[17] Y. Tenne and S. W. Armfield, "A versatile surrogate-assisted memetic algorithm for optimization of computationally expensive functions and its engineering applications," *Studies in Computational Intelligence*, vol. 92, pp. 43–72, 2008.

[18] S. Srisomporn and S. Bureerat, "Geometrical design of plate-fin heat sinks using hybridization of MOEA and RSM," *IEEE Transactions on Components and Packaging Technologies*, vol. 31, no. 2, pp. 351–360, 2008.

[19] S. Kanyakam and S. Bureerat, "Comparative performance of surrogate-assisted MOEAs for geometrical design of pin-fin heat sinks," *Applied Mathematics*, vol. 2012, Article ID 534783, 14 pages, 2012.

[20] A. Husain and K. Y. Kin, "Optimization of a microchannel heat sink with temperature dependent fluid properties," *Journal of Materials Processing Technology*, vol. 208, no. 1–3, pp. 499–506, 2008.

[21] J. Yoo and C.-Y. Lee, "Topology optimization of a swing arm type actuator using the response surface method," *Microsystem Technologies*, vol. 13, no. 1, pp. 21–31, 2007.

[22] W. Hu, L. G. Yao, and Z. Z. Hua, "Optimization of sheet metal forming processes by adaptive response surface based on intelligent sampling method," *Journal of Materials Processing Technology*, vol. 197, no. 1–3, pp. 77–88, 2008.

[23] N. Nariman-Zadeh, A. Darvizeh, A. Jamali, and A. Moeini, "Evolutionary design of generalized polynomial neural networks for modelling and prediction of explosive forming process," *Journal of Materials Processing Technology*, vol. 164-165, pp. 1561–1571, 2005.

[24] T. Goel, R. Vaidyanathan, R. T. Haftka, W. Shyy, N. V. Queipo, and K. Tucker, "Response surface approximation of Pareto optimal front in multi-objective optimization," *Computer Methods in Applied Mechanics and Engineering*, vol. 196, no. 4–6, pp. 879–893, 2007.

[25] Z. Zhou, Y. S. Ong, P. B. Nair, A. J. Keane, and K. Y. Lum, "Combining global and local surrogate models to accelerate evolutionary optimization," *IEEE Transactions on Systems, Man and Cybernetics Part C*, vol. 37, no. 1, pp. 66–76, 2007.

[26] Y. Jin, "A comprehensive survey of fitness approximation in evolutionary computation," *Soft Computing*, vol. 9, pp. 3–12, 2005.

[27] Y. S. Ong, P. B. Nair, A. J. Keane, and K. W. Wong, "Surrogate-assisted evolutionary optimization frameworks for high-fidelity engineering design problems," in *Knowledge Incorporation in Evolutionary Computation, Studies in Fuzziness and Soft Computing*, Y. Jin, Ed., pp. 307–332, Springer, Berlin, Germany, 2004.

[28] Y. Jin, "Surrogate-assisted evolutionary computation: recent advances and future challenges," *Swarm and Evolutionary Computation*, vol. 1, pp. 61–70, 2011.

[29] A. I. J. Forrester and A. J. Keane, "Recent advances in surrogate-based optimization," *Progress in Aerospace Sciences*, vol. 45, no. 1–3, pp. 50–79, 2009.

[30] K.-T. Chiang and F.-P. Chang, "Application of response surface methodology in the parametric optimization of a pin-fin type heat sink," *International Communications in Heat and Mass Transfer*, vol. 33, no. 7, pp. 836–845, 2006.

[31] Y. Tenne and S. W. Armfield, "A framework for memetic optimization using variable global and local surrogate models," *Soft Computing*, vol. 13, no. 8-9, pp. 781–793, 2009.

[32] P. Breitkopf, H. Naceur, A. Rassineux, and P. Villon, "Moving least squares response surface approximation: formulation and metal forming applications," *Computers and Structures*, vol. 83, no. 17-18, pp. 1411–1428, 2005.

[33] M. Xiao, B. Raghavan, P. Breitkopf, R. F. Coelho, C. Knopf-Lenoir, and P. Villon, "Surrogate models based on constrained POD for optimization," in *Proceedings of the 4th European Conference on Computational Mechanics (ECCM '10)*, Paris, France, May 2010.

[34] M. Xiao, P. Breitkopf, R. F. Coelho, C. Knopf-Lenoir, M. Sidorkiewicz, and P. Villon, "Model reduction by CPOD and Kriging: application to the shape optimization of an intake port," *Structural and Multidisciplinary Optimization*, vol. 41, no. 4, pp. 555–574, 2010.

[35] B. Raghavan, M. Hamdaoui, M. Xiao, P. Breitkopf, and P. Villon, "A bi-level meta-modeling approach for structural optimization using modified POD bases and diffuse approximation," *Computers and Structures*, 2013.

[36] B. Raghavan and P. Breitkopf, "Asynchronous evolutionary shape optimization based on high-quality surrogates: application to an air-conditioning duct," *Engineering with Computers*, 2012.

[37] E. A. Kontoleontos, V. G. Asouti, and K. C. Giannakoglou, "An asynchronous metamodel-assisted memetic algorithm for CFD-based shape optimization," *Engineering Optimization*, vol. 44, no. 2, pp. 157–173, 2012.

[38] H. Zhu, L. Liu, T. Long, and L. Peng, "A novel algorithm of maximin Latin hypercube design using successive local enumeration," *Engineering Optimization*, vol. 44, no. 5, pp. 551–564, 2012.

[39] S. D. Georgiou and S. Stylianou, "Block-circulant matrices for constructing optimal Latin hypercube designs," *Journal of Statistical Planning and Inference*, vol. 141, no. 5, pp. 1933–1943, 2011.

[40] J. M. Parr, A. J. Keane, A. I. J. Forrester, and C. M. E. Holden, "Infill sampling criteria for surrogate-based optimization with constraint handling," *Engineering Optimization*, vol. 44, no. 10, pp. 1147–1166, 2012.

[41] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: improving the strength pareto evolutionary algorithm for multiobjective optimization," in *Evolutionary Methods for Design, Optimization and Control*, pp. 95–100, 2002.

[42] S. N. Lophaven, H. B. Neilson, and J. Sondergaard, "DACE a MATLAB kriging toolbox," Tech. Rep. IMM-TR-2002-12, Technical University of Denmark, 2002.

[43] R. L. Harder and R. N. Desmarais, "Interpolation using surface splines," *Journal of Aircraft*, vol. 9, no. 2, pp. 189–191, 1972.

[44] M. E. Botkin, "Shape optimization with buckling and stress constraints," *AIAA Journal*, vol. 34, no. 2, pp. 423–425, 1996.

[45] K. Deb, A. Pratap, and T. Meyarivan, "Constrained test problems for multi-objective evolutionary optimization," in *Evolutionary Multi-Criterion Optimization (Zurich, 2001)*, vol. 1993 of *Lecture Notes in Computer Science*, pp. 284–298, Springer, Berlin, Germany, 2001.

[46] E. Mezura-Montes and C. A. Coello Coello, "Constraint-handling in nature-inspired numerical optimization: past, present and future," *Swarm and Evolutionary Computation*, vol. 1, no. 4, pp. 173–194, 2011.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 943702, 13 pages
http://dx.doi.org/10.1155/2013/943702

Research Article

A Digital Interface for the Part Designers and the Fixture Designers for a Reconfigurable Assembly System

Vishwa V. Kumar,[1] Salik R. Yadav,[2] F. W. Liou,[1] and S. N. Balakrishnan[3]

[1] *Department of Manufacturing Engineering, Missouri University of Science & Technology, Rolla, MO 65409, USA*
[2] *Health Care Service Corporation, Chicago, IL 60601, USA*
[3] *Department of Mechanical & Aerospace Engineering, Missouri University of Science & Technology, Rolla, MO 65409, USA*

Correspondence should be addressed to F. W. Liou; liou@mst.edu

Received 25 February 2013; Revised 7 June 2013; Accepted 7 June 2013

Academic Editor: Zhuming Bi

This paper presents a web-based framework for interfacing product designers and fixture designers to fetch the benefits of early supplier involvement (ESI) to a reconfigurable assembly system (RAS). The interfacing of the two members requires four steps, namely, collaboration chain, fixture supplier selection, knowledge share, and accommodation of service facilities so as to produce multiple products on a single assembly line. The interfacing not only provokes concurrency in the activities of product and fixture designer but also enables the assembly systems to tackle the spatial and generational variety. Among the four stages of interfacing, two steps are characterized by optimization issues, one from the product customer side and the other from the fixture designer side. To impart promptness in the optimization and hence the interaction, computationally economic tools are also presented in the paper for both of the supplier selection and fixture design optimization.

1. Introduction

In the current scenario, immense pressure faced by manufacturing firms to meet the precariously varying demand is a fait accompli. This is so because the market always demands a very large product variety among different products manufactured by any firm. The varying trend of the product variety in a product family is sought not only across different market segments (spatial variety) but also across different generations of the same product (generational variety) [1]. Obviously, these trends of product variety are valid for almost every kind of product development process, assembly systems not being the exception. Meeting the challenges posed by two kinds of variety is even higher in the case of assembly systems as it may require huge investments in different assembly lines or imparting reconfigurability to a single assembly line.

In multistation assembly processes, fixtures play the central role as most of the assembly faults are related to them. Moreover, to impart the desired product variety and to ensure unhampered product development of a product family being assembled on a single line, the assembly fixtures are required to possess certain sophistications. These sophistications include reconfigurability [2], stability [3], and fault diagnosability [4].

Reconfigurable assembly fixtures are those on which production of multiple parts can be carried out. That is, a fixture is sought to possess reconfigurability so as to produce multiple products in a single reconfigurable assembly system [2]. Most of the previous research on assembly fixtures is concerned with fault diagnosis [5–7] owing to the fact that most of the faults occurring in assembly parts are due to failures of locating pins, clamps, or supports of the fixture. This is so because during ramp up, launch fixtures are most prone to failures [8]. However, reconfigurability is also a topic worth paying attention as it can drastically alleviate the set-up and processing costs by processing multiple parts on the same fixture [9]. This helps the firm to realize the spatial variety at lower cost and at the same time enables the firm to face severe market diversity. Further, in order to reduce the overall product development time and hence to obtain higher market share, the tool designers (such as die or fixture designers) are severely affected. This is so because these tool designers decide the moment start of serial production [10]. A firm using assembly lines also desires to reduce the total number of tool designers involved in the production of the product. That is, a company might be using 10 to 20 fixtures for the production of a car body, but later it may come up with a strategy to reduce it to 5 or 10 fixtures [11]. Thus, in order to meet the quest for obtaining higher market share, globalization and distribution, it is necessary to develop cooperative supply chains with lesser lead times. Keeping in mind the globalization and importance of a supply chain, a firm might prefer to buy the assembly fixtures from the outside or inside fixture suppliers; those are basically fixture designers, so that the fixture suppliers design the assembly fixtures as per demanded by the product customers. In such case, it is beneficial for the firm to reduce the number of fixture suppliers as well to a small number of system specialists. Referring to the industrial contexts, Audi has reduced its tooling suppliers by more than half over four years [12]. Thus, to meet the market diversity and varying trends, it is necessary to develop a cooperative relationship between the fixture designers and the product customers. Research addressing the collaborative relationship between fixture designing and product development is essential in the context of a reconfigurable assembly system (RAS).

This paper addresses the problem of developing collaborated fixture design and product development processes by adopting a web-based framework. The developed web-based framework presents a strategy to impart a cooperative approach to concurrently perform fixture design and assembly line production. The paper opens a way to integrate the process chains of the product development and fixture designers through the web for a RAS. The web-based framework presented hereby is termed as cyberdesigning of RASs.

Traditionally, the product development in assembly system and fixture making is resolved separately. However, for developing a product family on a single assembly line the fixture design must be in cooperation with product design; enabling a fixture to be suitable for multiple part models needs reconfigurability of the fixture [13, 14]. The collaboration between product and fixture designers enables the assembly system to excel in the longer run because locating and clamping elements in the fixtures and their positions depend upon the work-piece design [15, 16]. Recently, concurrent engineering has emerged to be a great tool for reviewing and coordinating different aspects in the product life cycle. Several researchers have laid emphasis on the coordination requirements in product development. However, most of their work is directed towards coordinating activities: machining processes [17, 18], stamping product development [10], and sheet metal forming [19]. Though few researchers have proposed that activities in product development and tool designing can be coordinated through intelligent systems [20], computer aided design [21], and so forth, the coordination of activities with reference to the context of product design and fixture design remains sparsely attended.

Recently, a few researchers have advocated the use of the internet in coordination of various activities thorough knowledge exchange, e-bidding, online communication, and so forth, thus making the internet more than an information display system. For example, product design and manufacture roles can be resolved using web technology [22–24]. Thus, it is established in the literature that web technology can be a useful tool for carrying out concurrent and intelligent decision making to provoke early supplier involvement.

Utilizing the utility of web technology, this paper presents a web-based interfacing strategy to enable an assembly system to (1) efficiently take care of the desired spatial and generational variety in the product family developed in an assembly system; (2) impart reconfigurability to the assembly system so as to produce multiple products on a single line; (3) receive uninterrupted assistance from fixture designers at various level of product development; (4) provide a tool for optimal supplier selection in stipulated time frame; (5) create a web interface for mandatory knowledge share between the assembly system and fixture designer; and (6) present a heuristic for optimizing fixture design in stipulated time frames.

The rest of the paper is organized as follows. Section 2 explains various quests, scopes, and advantages of coordinating product development in assembly system and fixture design. It is followed by the web-based interfacing framework for product customer and fixture designers in Section 3. Later, Section 4 illustrates the optimization issues in the web-based interaction. Finally, the paper is concluded in Section 5.

2. Inspecting Collaborated Fixture Design and Product Development Processes

2.1. Integration of Fixture Design and Product Development. The overall process chain for developing a product on an assembly comprises of product design, assembly process planning, fixture design, and manufacturing planning. The design of the assembly fixture determines whether it can support multiple parts or not. For any assembly line, the product development activities and fixture design are closely related, and in the case of RASs, it is even more crucial [2]. As changing characteristics of parts alters the design of fixtures, so the reconfigurability of assembly systems mainly depends upon the reconfigurability of the fixtures. Thus, the link between the product development and fixture design is in fact the link between product design and serial production. Figure 1 portrays the interrelationship between the activities of product design and fixture design.

In this paper, it is assumed that the fixture supplier realizes the design of the fixtures as per the demands of the product customer. The fixture supplier cum designer will be generically called fixture designer throughout this paper. In the reconfigurable assembly system, sequential order processing is carried out where finished and prepared documents are formerly transferred to an internal or external fixture designer after the completion of the product development. In fact, the fixture designer has to design and build a manufacturing resource without perturbing the

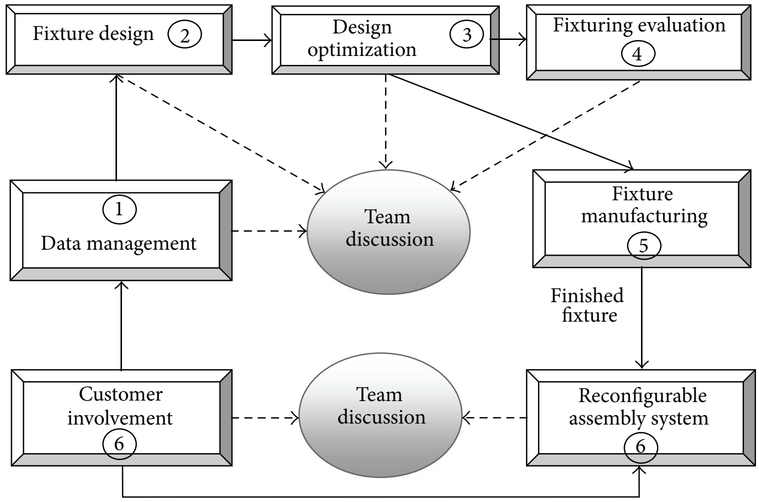


Figure 1: Relationship between product development and fixture development.

fixture-based design of the product. This aspect portrays that a missing communication between the product and fixture development may cause various inconsistencies. These inconsistencies include poor assembling properties of the part in spite of the functional ability, expensive assembling of the part, and suitability of the fixture for multiple parts, and so forth.

The integrated fixture design and product development can reduce the development cycle time. This is possible as performing preliminary designing of the assembly part and selection of assembly fixtures concurrently instead of waiting for the complete design of the assembly part can save an enormous amount of time. Moreover, improvement in development quality also becomes viable due to consideration of potential quality problems at earlier design stages. Though integrated fixturing and product development are promising, it requires very close cooperation between part assembler and fixture designer. Figure 2 portrays integration schema for the product and fixture design.

2.2. Synergizing Relationship between Product Customer and Fixture Designer. Considering the importance of communication between reconfigurable assembly systems and the external or internal fixture designers, the information regarding the parts to be assembled on the fixture, such as number of parts and tooling configuration of each part, are crucial for both the product developers and the fixture designers. Once this information is received by the fixture designers, they utilize it to design the assembly fixtures. However, they also tend to optimize the design of the assembly fixtures, as discussed in the next section, which further increases the criticality of the cooperation. Moreover, the assembly process carried out in an RAS is a complex process requiring large interaction between assembly process variables such as deciding assembly robots, welding guns, and grouping of tasks. This is so because in RASs, two similar parts (say car windows for two different car models) may require an assembly robot while the other may require a welding gun for a particular process. This and other implications of RASs imply that developing products on reconfigurable assembly lines requires extensive knowledge of various processes, their capabilities, and limitations. That is, it is difficult to impart reconfigurability by a product designer without comprising with the product development processes, because of extensive requirement of knowledge and experience required, which is very unlikely in practical situations. At the same time, the fixture designers or fixture experts (designer cum suppliers) often use heuristics based on extensive experience and knowledge and can ultimately assist to carry out product development of a product family on an assembly line in a relatively short time. Hence, the cooperation between the fixture designers and the product development processes for an RAS can lead to critical savings in the product development time.

Making the fixture designers an arranging partner in the entire process chain of product development can result in efficient integration of product development and fixture development, if partnered with an early and active involvement. This scope can minimize the cost, time, and design faults leading to advantageous situations for both the customer and the fixture designers. Several researchers support this aspect; particularly Tang [11] and Eversheim [12] have proposed the active involvement of die makers in the case that stamping product development can lead to much improved results, thus, making the active cooperation between fixture designers and product development in RASs even more promising. Early integration of the customer process may lead to an earlier start of the fixture designing activities as well as a shift between the activities in the activities processed. However, an early integration is not an easy job in the context of global manufacturing, that is, when the fixture designers are geographically dispersed at large distances. In such cases, the cooperation between the fixture designers (designers cum suppliers) with the product designers and hence the concurrent performance of preliminary product design processes and fixture designer selection can be accomplished using a web-based environment. This paper presents a web-based framework termed “cyberdesigning of RASs” for cooperating

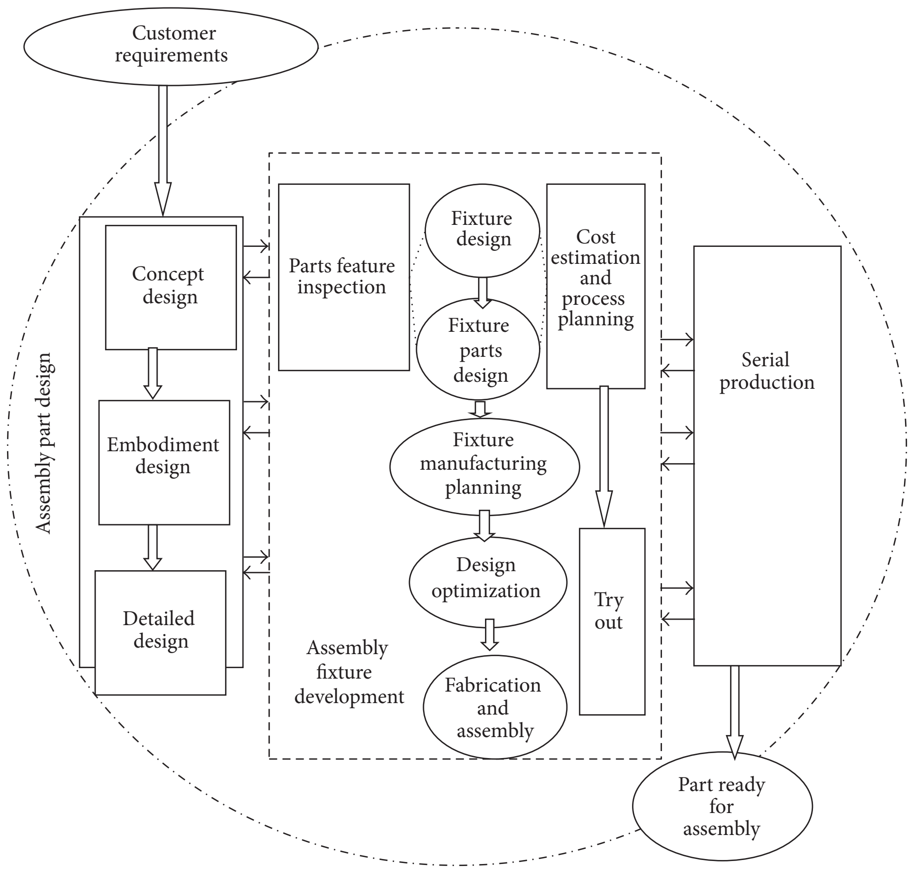


Figure 2: Integrated assembly fixture and part development.

and integrating the activities of fixture design and product development in an RAS as discussed in the next section.

3. Cyberdesigning of RASs

The web-based framework "cyberdesigning of RASs" is developed to conceptualize a cooperative integration of fixture designers with the product development of a product family on a single assembly line. In this framework, the activities in fixture design or the activities of fixture designers are juxtaposed with product development activities regarding the product family on the assembly line. This juxtaposition is accomplished by interfacing the product customer and fixture designer through the framework "cyberdesigning of RASs" as shown in Figure 3.

This framework performs the following functions for interfacing the product customer's and fixture designer's activities.

(a) *Collaboration chain*: The collaboration between product customer and fixture designer can be done using a bill-of-materials-based partnership chain.

(b) *Fixture designer selection*: A fixture designer (designer cum supplier) selection module can be developed to select appropriate fixture designers.

(c) *Information sharing*: Information sharing includes facilities for know-how, data, and experience exchange between the RAS and the fixture designers.

(d) *Service facilities*: The cyberenvironment for fixture design and product assembly can be well supported by providing different functions such as user login management, online information exchange, and website administration.

All of these functions together constitute the job accomplished by the web-based interface "cyberdesigning of RASs." The descriptions of the different tasks of this interface are discussed in the following subsections.

3.1. Collaboration Chain. The collaboration chain can be defined as a partial supply chain model, consisting of basic interface between the fixture designer and the product customer. This two-stage supply chain can be represented using the notion of the generic bill of material (GBOM) of the involved product family. Though the market desires a very diverse product family, huge savings in manufacturing costs can be obtained via commonalizing some parts and differentiating others. For example, in a car body assembly system, the manufacturer can choose to differentiate the part "car bonnet" for different market segments, whereas it can offer the same part "car window" to each of the market

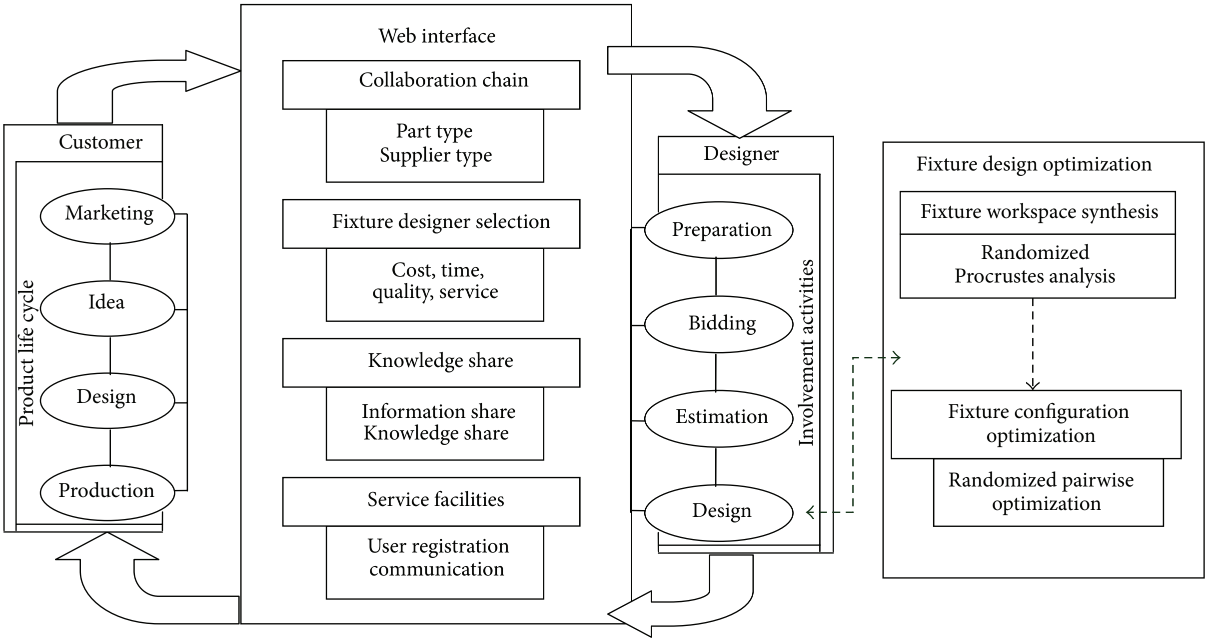


Figure 3: Prototype of web-based interface for product customer and fixture designer.

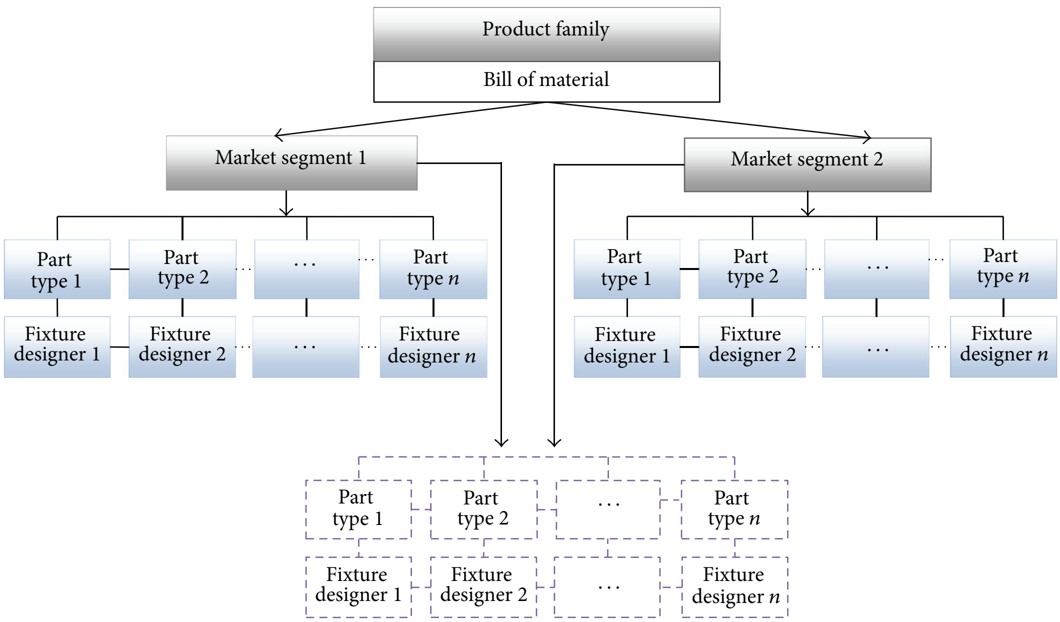


Figure 4: A GBOM-based collaboration chain: dotted boxes represent common parts, and solid ones indicate differentiated parts across two-market segments.

segments. Consequently, the need for the reconfigurable assembly fixture is due with the differentiated part types only.

Figure 4 portrays a GBOM employed for the collaboration chain process. In the collaboration chain process, first the customer requirements are obtained and the members of product family are planned. Then, reconfigurable and nonreconfigurable fixtures are decided as per the common or different part models to be assembled on a fixture. This is followed by the selection of individual fixture designers. For this purpose, web-based frameworks such as WeBid exist in the literature which uses BOM-based product models to ensure early supplier involvement [25, 26]. Though BOM-based collaboration chains can bring about early supplier involvement in general product development, in case of reconfigurable assembly systems it is very complex. This is because of the very high level of responsibility and execution for technical and organizational coordination with the customers. In such case, selection of adequate fixture and tool suppliers requires advertisement for each tool or fixture. Moreover, the amendments in initially planned schedule may

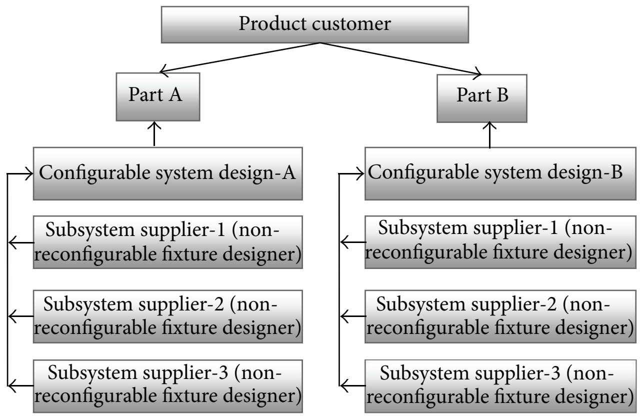


FIGURE 5: Conceptualization of virtual supplier for building multi-level collaboration chain.

raise the costs in case of multiple fixtures considered at a time, thus making the job of customer cumbersome as the number of fixture increases.

Moreover, the numbers of fixtures or tool suppliers are desired to be reduced, particularly in the automobile industry [10], so as to alleviate the costs incurred in the procurement of fixtures and other tools. An endeavour in this direction inspired from Eversheim [12] can be brought about by decreasing the number of suppliers in direct communication with the product manufacturer. This can be done by involving intermediate fixture and tool suppliers.

This paper utilizes the notions of system supplier, reconfigurable fixture designer cum supplier, and nonreconfigurable fixture designers cum supplier together forming a virtual supplier as shown in Figure 5. A system supplier is a reconfigurable assembly fixture designer meant for differentiated part types across the product family, whereas subsystem suppliers are the nonreconfigurable fixture designers meant for common part types across the product family. The subsystem suppliers together with the system suppliers constitute a virtual supplier by simultaneously contributing in the complete fixture development. It results in two improvements in the product manufacturer-fixture designer relationship. These improvements include reduction in the number of directly interacting fixture designers and transfer of overall job of fixture designing from the product manufacturer to the system supplier.

3.2. Fixture Designer Selection. For early supplier involvement, the selection of the right fixture designer is a critical issue. The criticality in the fixture designer selection lies in that a set of conflicting criteria are to be considered to realise the effective selection. These criteria range mainly in the form of technical requests along the process chain of development, fixture design and tool production, assembly, optimization, and so on. The fixture designer selection problem addressed in this paper is a multiple supplier selection problem to obtain an optimal set of fixture designers in the multilevel collaboration chain. The fixtures to be used in RAS are classified, as mentioned earlier, into reconfigurable and nonreconfigurable fixtures required, respectively, for differentiated and common part types such as components of car bodies. The two kinds of fixtures are supplied, respectively, by system and subsystem supplier. The complexity of this problem may vary from problem instance to problem instance as there may be several alternative fixture designers for the same fixture or one fixture designer may develop more than one fixture. Nonetheless, the fixture maker selection problem in this paper is concerned with the selection of the sequence of the suppliers.

The formulation of the fixture designer selection problem is inspired form Gupta and Nagi [27], Sha and Che [28], and Lin and Wang [29]. The criteria considered in the supplier selection problem can be cost, lead time, service level, quality, and so forth. The contribution of these criteria with respect to the optimality in the supplier sequence depends upon the weight or priority assigned to these criteria. They proposed heuristics based on explicit enumeration which can be used to find the optimal combination of system and subsystem assembly fixture suppliers.

Let S_{jk}^{l} be the set of parts for which the lth fixture designer develops the fixture at the plant j^{l} for fixture type k^{l}. Fixture type means reconfigurable, nonreconfigurable, large size, medium size, or small size fixtures. Then, the fixture maker selection problem can be formulated as

$$\begin{aligned} \text{Maximize} \quad & \sum_{l=1}^{n} \left[w_1 \cdot C_{jk}^{i}\left(S_{jk}^{l}\right) + w_2 \cdot \Gamma_{jk}^{i}\left(S_{jk}^{l}\right) \right. \\ & \left. + w_3 \cdot Q_{jk}^{i}\left(S_{jk}^{l}\right) + \cdots \right] \end{aligned} \tag{1}$$

Subject to

$$\begin{gathered} \left(S_{jk}^{l}\right) \cap \left(S_{jk}^{m}\right) = \varphi, \\ \operatorname{Card}\left(\bigcup_{l=1}^{n(V)} \left(S_{jk}^{l}\right) \right), \end{gathered} \tag{2}$$

where w_1, w_2, and w_3 are the priorities or weights of the different criteria considered in the objective function. Solving the above formulated problem means finding a feasible combination of fixture designer defined by

$$\left\langle \left(j^{l}, k^{l}, S_{jk}^{l} \right) \right\rangle_{l=1}^{n(V)} \tag{3}$$

with "n" being the total number of the of fixture designers in the Vth combination. $C_{jk}^{i}(S_{jk}^{l})$, $\Gamma_{jk}^{i}(S_{jk}^{l})$ and $Q_{jk}^{i}(S_{jk}^{l})$ are the ranks of the lth fixture designer with respect to cost, lead time, and quality, respectively. For enumerative purposes, each of the criteria, whether qualitative or quantitative, needs to be quantified. For this purpose, a normalized rank or rating is used to convert the quantitative or qualitative criteria as follows:

$$R = \begin{cases} \dfrac{(V - \mathrm{LL})}{(\mathrm{UL} - \mathrm{LL})} & \text{criteria are to be maximised} \\[2ex] \dfrac{(\mathrm{UL} - V)}{(\mathrm{UL} - \mathrm{LL})} & \text{criteria are to be minimised,} \end{cases} \tag{4}$$

Table 1: Brief sketch of knowledge share components between part designer and fixture designer.

Part design activities	Fixture design activities				
	Material cost estimation	Fixture cost estimation	Process planning	Fixture design optimization	Fixture model design
Requirements of each market segment					
Commonality/differentiation decision					
Material selection	X	X			
Concept generation	X				
Individual parts configuration		X		X	X
Tooling elements and clamps characterization	X	X		X	X
Dimension specification		X	X		X
Generation of part/parts for assembly					X

where R is the normalized rank or rating and V is the estimated value of the quantitative or qualitative criteria converted to a quantitative value. UL is the upper limit and LL is the lower limit.

3.3. Knowledge Sharing

3.3.1. Need for Knowledge Sharing. For effective realization of early supplier involvement in the process chain, the knowledge share between the product designer and the fixture maker is the key factor. The term knowledge sharing includes both the design knowledge as well as design information sharing. The product development cycle comprises of different steps such as conceptual design steps, embodied design steps, and detailed design stage. Table 1 reveals knowledge sharing components between the part designer and fixture designer. The emphasis utilizing ESI should be to have fixture designer involvement at as many steps as possible. During the conceptual design steps, the lack of exhaustive product information may trigger the fixture designer to hesitate to specify all the workholding devices and tools that are required, thus making it difficult to perform exact fixture cost calculation. The fixture designers can help at different stages of product development cycle in the following way.

In the proposed web-based framework,

(a) the customers' verbal instructions are utilized in making a conceptual sketch of or model to generate concept parts (differentiated or common) and material selection. The involvement of fixture designers at this stage can help in identifying the most suitable material as per the sheet metal properties, forming qualities, and cost requirements.

(b) At the embodiment design stage, that is, during the part configuration processes, the assessment of fixture's cost, part's fastening properties, and utilization of previously made sketches can be well aided by the fixture designers. The cost evaluation of the previously designed part sketch is the most crucial part which is done on the configuration level.

(c) During the detailed design stage, part shape refinement, dimension specification, tolerance specifications, and clamping or tooling location specification is done. The part geometry is the most important aspect to be attended to from the side of the part designer. This is because, in assembly systems, the geometry features determine not only assembling properties but also the functional requirements of the part. For example, holes in the parts are used for different purposes such as fastening, guiding and aligning, or reducing the weight of the part. For fixture designers, part design determines the configuration of the assembly fixture designed by them. In the case of system suppliers, they have to make reconfigurable fixtures for which they require tooling or clamping specifications. They tend to minimize the fixture workspace envelope (FWE) for which they may use several alternative heuristics. The minimization of FWE during the fixture designing not only saves costs incurred during physical development of the fixture but also takes lesser space in the assembly systems. Moreover, it is also sought that the part design should be functionally acceptable as well as compatible with selected assembly processes so as to yield lower cost, shorter lead times, and higher quality.

The achievement of aforementioned objectives requires the consideration of the fixture type, number of parts assembled on a fixture, number of fixtures, and fixture development cost with respect to the part feature traits such as feature form complexity, number of tool locating elements, size, tolerance, and so forth. This requires a sound knowledge sharing between the product designer and the fixture designer. The proposed framework strives to resolve this issue by adopting any of the two knowledge sharing techniques as discussed below.

3.3.2. Mode of Knowledge Sharing. After selection of the right fixture designer for the assembly system, the early supplier involvement gets initiated from the organizational point of view. What remains yet to be materialized is to create a cooperative environment between the product designer and fixture designer. The next step in this direction, in the cyberdesigning of RASs, which follows the fixture supplier selection, is the knowledge share the quest for which is mentioned in the previous subsection. The knowledge share

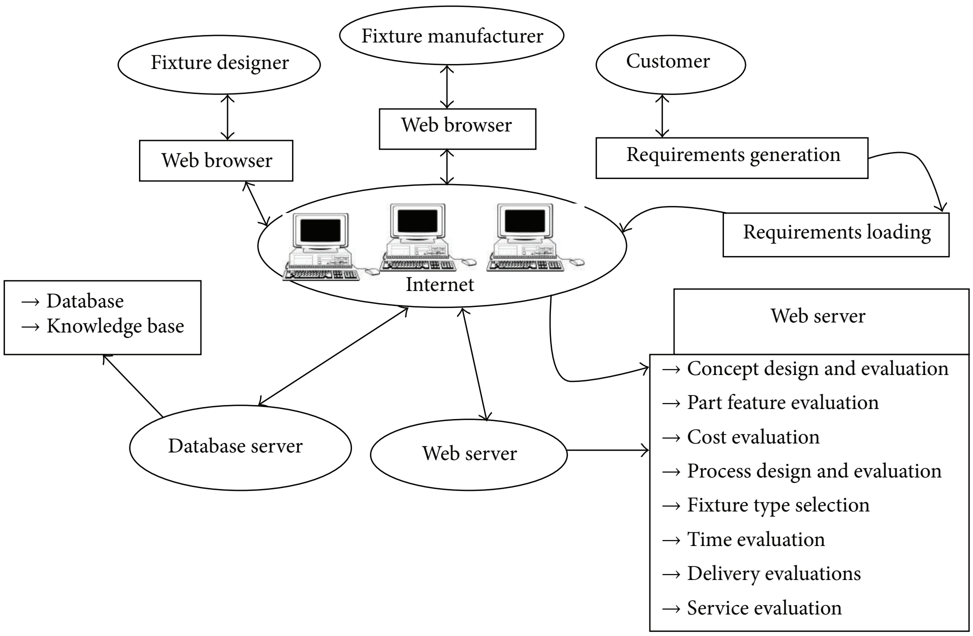


Figure 6: The knowledge share through web-based services in reconfigurable assembly line.

between the fixture designer and the product designer can be accomplished by either using knowledge bases furnished by fixture designers or by availing an online communication environment for messaging and data exchange between product designers and fixture designers. The former mode of knowledge share is called asynchronous knowledge share, while the latter one is known as synchronous knowledge share.

In the asynchronous knowledge share, the fixture designer provides its knowledge in web-based knowledge bases. The knowledge base can contain knowledge such as rules, case studies, heuristics, and demonstrations. Moreover, web-based design services can also be availed through the knowledge for concept design and evaluation, part design evaluation, assembly process design and evaluation, fixture set selection, and so forth, for analyzing different stages in developing the product family at the reconfigurable assembly line. Several design problems can be handled using the web-based services equipped with the knowledge base of the fixture designers. During part geometry design, the product designer can receive various options and consequences for cost, processes, and fixture design for reconfigurable as well as nonreconfigurable fixtures. These design services can be utilized to integrate the fixture designer with product development on a reconfigurable assembly system. High level information about knowledge object known as the metadata can be used to perform knowledge mining, searching and navigation. The search and navigation can be performed by adding keywords to the metadata so that specific searches can also be done. The user can browse through the keywords and explore into the knowledge base. For this purpose, Java Applet can be utilized to provide an interface for realizing the knowledge share between the product designer and the fixture designer which is clearly a future scope of research. The mechanism of asynchronous knowledge share is shown in Figure 6.

In synchronous knowledge share, an agent communication platform based on Java Agent Template Lite [30] can be built to enable online communication between the product designer and the fixture designer. A basic infrastructure is fetched by JATLite where an agent registration is done with agent registration route (AMR) by the way of name and passwords. The agent also transfers files and invokes other programs or actions on the computer. Moreover, it can also accomplish the tasks of wrapping existing part and fixture design programmes by enabling an automatic communication with other programs via a front end. This research recommends that while utilizing synchronous knowledge share between the product designers for RASs and the corresponding fixture designers, the construction of agents can be facilitated by JATLite. The agent can send and receive knowledge query and manipulation language (KQML) and extensible markup language (XML) languages by sending a KQML message with XML contents to the receiving agent for its main applications. The receiving agent performs required operations according to the content of the message. It does so by interpreting the XML-based message content followed by its use of the related data for system application so as to minimize the data communication errors.

The part designers, via an agent, can use the features of the differentiated parts or the common parts across the product family as design units to be used by the fixture designers and configuration and evaluation units in the fixture design. The part features such as number of tooling elements or tool locators and parts' clamping configurations can be defined through appropriate feature templates. Meanwhile, XML document can be used to describe the part features. The fixture designer through an agent can read and analyse the part feature information via KQML/XML parser. After analyzing the part features, the fixture designer evaluates the part model for evaluation at design stages. The evaluations include reconfigurable evaluation, cost and space evaluation of the part design and hence the fixture. If the part model contains any design flaw, then the part designer is informed through the part design agent. Otherwise, the fixture designer starts developing the assembly fixture. The fixture designer will try to optimize the FWE of the fixture in case of reconfigurable as well as nonreconfigurable fixtures so as to minimize the fixture design and development costs incurred on his side. The optimization of fixture design not only fetches benefit to the fixture designer but also to the product development processes in the RAS. This is so because better space utilization of space in the assembly system and economic fixture will be availed for the RAS. Few heuristics exist in the literature, for optimizing fixture design. This paper also incorporates optimization aspects for fixture design as per the parts' design. Extended heuristics are presented in this paper that can be used by the fixture designer once the part design information is obtained, evaluated, and approved by the fixture designers. The next subsection discusses the optimization aspect of fixture development.

Thus, the cyberdesigning of RASs contains two optimization aspects, one prior to the knowledge share that is the fixture design selection and the other posterior to the knowledge share that is the fixture design optimization as per the requirements of parts' design. These optimization aspects are discussed with adopted heuristics and examples in the next section.

4. Optimization Issues in Cyberdesigning of RASs

As discussed in the previous section, the function modules of "cyberdesigning of RASs" include collaboration chain, fixture designer selection, knowledge share, service facilities, and supplementing tools. The functions of service facilities include aiding users' registrations who are nobody but the product customer and fixture designer. The service facilities are planned so that the customer/user can view the information of the supplier but cannot manipulate it; likewise is true for the fixture designer. The supplementing tools such as JATLite agents and KQML and XML languages are used to accomplish the function of online interaction between product customer and fixture designer through discussion, messaging, information upload and download, and so forth.

Among the above recited functions of "cyberdesigning of RASs," the first function module, namely, collaboration chain characterization, does not ask severely for optimization. Rather, it involves the provision of developing a skeleton of collaboration chain by the customers during the earlier interaction between the customer and the designer. GBOM representation of the product family involved can be used to present a graphical navigation of the collaboration skeleton so that it provides a crucial deciding point not only to the customer but also to the fixture designers. Via the help of the products GBOM, fixturing bids are invited by the customer and are evaluated by the fixture designer. The evaluation by the fixture designer involves capturing of the customer requirements and identification of his own role in the collaboration. The customer requirements include the description of parts such as number of tool locators, clamps, and supporters for the part. The latter part of evaluation refers to realization of the fact that whether the fixture designer can play the role of a system or subsystem supplier depending on differentiated or common part, and hence reconfigurable or nonreconfigurable fixtures are to be developed. After the characterization of collaboration chain by the customer, different fixture designers submit their bids. The evaluation of bids of the entire potential fixture design by the customer is next realized.

Thereafter, supplier selection from the potential fixture designers is to be done. This requires the optimization of different conflicting criteria for which sophisticated techniques and heuristics are needed, particularly for large size problems. This optimization issue is discussed later.

4.1. *Prior-to-Knowledge Share Optimization.*

As discussed in the previous section, several criteria compete for consideration while accomplishing the supplier selection in an optimal way (1). Previously, the multiple-supplier selection problem, similar to our context, has been solved using explicit enumeration techniques. However, in cases involving a very large number of suppliers to be selected, it may be a cumbersome job to find the optimal solution in computationally feasible time. This paper proposes a random search metaheuristics, namely, artificial immune system (AIS) which is discussed below.

4.1.1. *Artificial Immune System.*

Despite AIS's recent emergence, it has well established itself as a potential optimization algorithm backed by superior results it has yielded [31]. AIS incorporates clonal selection and evolutionary principles altogether so as to impart learning among individuals, or in computational terms, the random solutions [32].

It is a population-based strategy which starts with the maintenance of a population P of antibodies. Each antibody is a random solution string encoding the sequence of feasible suppliers as given by (3). Obviously, the representation schema is a permutation schema encoding the index of the selected supplier and arranged in the sequence.

This population of antibodies undergoes cloning wherein each individual is cloned or copied based on its antigenic affinity (or objective function given by (1)). That is, more clones are generated for antibodies with higher affinity and vice versa. The antigenic affinity refers precisely to the string's fitness in terms of objective function or constraint satisfaction.

Thereafter, all the clones have undergone a maturation process through hypermutation whereby the stochastic rate of hypermutation is made dependent on the fitness. Precisely, rate of hypermutation is kept higher for clones with lower fitness and vice versa. The difference between hypermutation (AIS) and mutation (GA) lies in the fitness dependency of the mutation rate in case of AIS. From the hypermutated clones population, P, best individuals are selected stochastically and the whole procedure of cloning and hypermutation is iterated. AIS is shown to accomplish a search with a fair trade off between exploration and exploitation [31].

4.1.2. Implementation over a Case Study. An implementation of artificial immune system over a problem taken from Tsai [33] and converted to our context (the suppliers being fixture designer cum supplier) is presented in this section. The criteria considered in the problem are cost, lead time, delivery ability, service, and flexibility so that the objective function is given by

$$\begin{aligned} \text{Maximize} \quad a_v = \sum_{l=1}^{n} \Big[& w_1 \cdot C^i_{jk}\left(S^l_{jk}\right) \\ & + w_2 \cdot \Gamma^i_{jk}\left(S^l_{jk}\right) + w_3 \cdot Q^i_{jk}\left(S^l_{jk}\right) \\ & + w_4 \cdot D^i_{jk}\left(S^l_{jk}\right) + w_5 \cdot \rho^i_{jk}\left(S^l_{jk}\right) \\ & + w_3 \cdot F^i_{jk}\left(S^l_{jk}\right)\Big], \end{aligned} \tag{5}$$

where quantity represented by (5) is the assessment index of the vth sequence of the fixture designers cum suppliers. C^i_{jk} is the cost rating, Γ^i_{jk} is the lead time rating, Q^i_{jk} is the quality rating, ρ^i_{jk} is the service rating, and F^i_{jk} represents the flexibility rating of the lth supplier at plant "j^l" and fixture type "k^l".

Table 2 shows mandatory fixture designs required for parts of a car body panel. The search space for this problem can be represented by a search tree. The weights considered hereby are taken to be 0.449, 0.133, 0.129, 0.028, 0.203, and 0.059, respectively, for cost, lead time, quality, delivery, service, and flexibility. The solution space for the problem can be represented via a search tree, and the optimal solution for this problem varies from combination to combination of the weight of the six criteria. For the abovementioned AIS algorithm is coded over this problem on C++ compiler and run till the optimal result is obtained. The optimal sequence found is $(1, 1, \{\text{LEP 01}, \text{LEP 02}\})_{1\text{-}}$ $(6, 2, \{\text{MIP 04}\})_{2\text{-}}$ $(2, 3, \{\text{SIP 03}, \text{SIP 05}\})_{3}$.

The solution for this problem is found after 100 iterations and is more promising on large data sets as compared to explicit enumeration in terms of computational time.

4.2. Posterior-to-Knowledge Share Optimization. Once the fixture designer is selected and part design, fixture designer capabilities, and so forth get realized through knowledge share, the selected fixture designers optimize the design of the fixture as per the parts associated with the fixture. Information like part configuration (tooling elements, tool locators) is utilized for this purpose. However, in order to reduce fixture development cost, fixture workspace envelope, and so forth, the fixture designer tends to optimize the fixture design. With reference to speedy and compact interaction to enable web-based cooperation, the viability of the heuristics to yield optimal fixture design used by the fixture designer in a stipulated time frame is a crucial issue. This paper extends

Table 2: Fixture designer alternatives for a car body panel.

Part	Fixture configuration type	Fixture designer options
LEP 01	1 (reconfigurable fixture)	1, 3
LEP 01	1 (reconfigurable fixture)	1, 6
SIP 03	3 (nonreconfigurable fixture)	2, 4, 6
MIP 04	2 (nonreconfigurable fixture)	5, 6, 7
SIP 05	3 (nonreconfigurable fixture)	2, 8, 9

a deterministic heuristic to optimize fixture configuration that can be used by the fixture designer via agent or other software to yield quality solutions in real time. The following subsection presents the extended stochastic metaheuristics for obtaining a near optimal/optimal fixture design.

4.2.1. Stochastic Procrustes-Based Pairwise Optimization. Lee et al. [34] has presented a genetic algorithm-based workspace synthesis analysis for fixturing of a stamping part family. They proposed to minimize the space required by each locator to produce a part family and thereby synthesizing the single fixture workspace. Though their approach substantially reduces the fixture workspace an analytical elimination of unlikely sets of solutions is missing that may impart swiftness to the workspace minimization. Moreover, the number of variables involved in minimization is directly proportional to the number of locators and parts causing the algorithm to consume very large amounts of time.

Kong and Ceglarek [2] proposed a methodology containing Procrustes and pairwise optimization. They utilized Procrustes analysis for workspace synthesis by obtaining best superposition of different configurations corresponding to different parts. The Procrustes analysis uses a least squares method with isotropic scaling, rigid translation, and rotation. It accomplishes preliminary fixture configuration design by best matching one configuration (set of points with multi-dimensions) to the others. For multiple parts, generalized Procrustes analysis is applied where different pairs from N number of parts (each with k number of tooling elements) are formed and superimposed to each other through translation and rotation in a Cartesian coordinate system to minimise the residual error or accumulative difference between two fixture configurations in the pair. In order to speed up the optimization, pairing is done such that a pair comprises of a part and the centroid of the remaining $N - 1$ parts.

Once the optimal superimposition of different configurations is obtained, the diameter of fixture workspace envelope (FWE) is determined by minimizing the diameter of the largest circle (out of k circles) containing N tooling elements (locating points) corresponding to each part. The diameter of the largest circle is minimized through making N pairs, each with one locating point and centroid of the rest of the $N-1$ locating points and conducting search optimization for this configuration pair to minimize the diameter of the largest circle (say ith one).

The objective function for the fixture configuration optimization is given by

$$\text{Min}\left(\text{Max } C_i\right) \quad 1 \le i \le k. \tag{6}$$

Though this deterministic Procrustes-based pairwise optimization achieves the optimization of fixture design, the absence of proper ordering of pair formation in each iteration impacts the computational time severely. An unplanned ordering strategy for each search optimization to compress the two members of the pair would result in the exact optimal solution at the cost of computational time. In order to impart further momentum to early supplier involvement, the computational time should be given priority over exact optimality of the solution. For this purpose, this paper randomizes the search procedure as follows.

(a) During each iteration of Procrustes analysis, the pair formation is generated by randomly selecting a locating point and taking the centroid of $N-1$ locating points. It is ensured, in each of the iterations, that the point which is once selected should not be selected again. Once N pair formations is followed by sum of squares of each pair of configuration, one iteration is completed. Repeating Procrustes analysis in this way yields a near optimal/optimal superimposition of part configuration.

The sum of squares of each pair of configuration [2] is given by

$$D(X_1, X_2, \dots, X_n,) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left\|(X_i R_i + l_k T_i) - (X_j R_j + l_k T_j)\right\|^2, \tag{7}$$

where X_i is the configuration matrix, R_i and T_i are the rotation and translation vectors, and l_k is the matrix with all the tooling elements of a part configuration.

(b) After determination of the superimposition of the part configuration, the diameter of the largest circle containing N parts is minimized. This is accomplished by pairwise optimization [2] as told above. Computational economy with respect to time can be brought about to this procedure by stochastically selecting a part while doing any pair formation in each of the iterations. This can be done by calculating the distance of locating elements from the centre of the circle and normalizing them. The normalized distance of a locating point is used in deciding selection of a point for the pair formation with the rest of $N-1$ points. That is, a point with lesser value normalized distance is selected first for pair formation and that with higher value is selected later.

This extends the Procrustes-based pairwise optimization by imparting to it a stochastic outlook. This helps in obtaining a sub optimal solution in a stipulated time frame.

The extended Procrustes-based pairwise optimization is implemented over an example taken from Kong and Ceglarek [2], and the stopping criterion for the algorithm is kept as the maximum number of generation. Different trials are preformed, and the diameter obtained versus trial number plot is drawn as is shown in Figure 7.

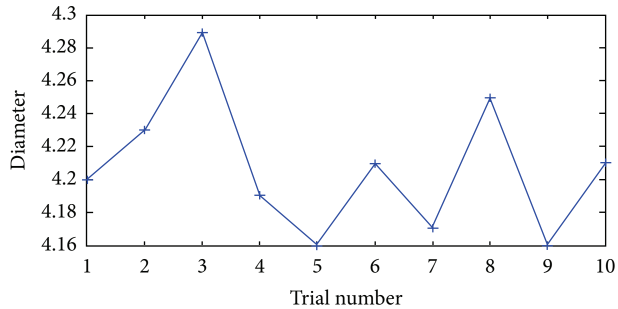


Figure 7: Performance of stochastic Procrustes-based pairwise optimization for stipulated number of iterations or time frame.

As reflected from the figure, for most of the trials the diameter of the largest circle is between what is found by Lee et al.'s approach or Kong and Ceglarek approach. Moreover, only once the diameter value found by stochastic Procrustes-based pairwise optimization is equal to 4.29 otherwise it is near to 4.16 units. Thus, the stochastic Procrustes-based pairwise optimization algorithm results in a viable solution in stipulated time.

5. Conclusions and Future Research

This paper attempts to conceptualize the product development of a product family in a reconfigurable assembly system in which product development is performed concurrently with fixture design. The fixture design is accomplished as per the parts' requirements by internal or external suppliers, and active interaction is simulated through a web-based interface so as to have early supplier involvement. The proposed web-based interface involves four steps of interaction where two of them are concerned with optimization issues. The first optimization problem is related to fixture designer selection by the product designers, while the latter one is optimization of fixture design by the fixture designer. To simulate promptness to the interaction between the two members, both of the optimizations can be aided with stochastic optimization algorithm.

Though the cooperation and integration of the product designer and fixture designer activities promise to maximize the market share, this aspect is not analyzed explicitly by this paper, thus being a future scope of research. Moreover, agents can be utilized more compactly to imbue sophistications and intelligence in the communication.

As synchronous knowledge share, an agent communication platform based on Java Agent Template Lite is built to enable online communication between the product designer and the fixture designer, and being used by a user group without any complain so far, the efficacy of the proposed digital interface can be established. At the same time the multiple-supplier selection problem is also demonstrated to be near optimally solved using a sophisticated algorithm in realistic computational time. Holistically, we are encompassing both of the products dependent and nonproduct dependent part in general. A more discreet and decisive model for reconfigurable or nonreconfigurable models is an area of further research.

References

[1] M. V. Martin and K. Ishii, "Design for variety: developing standardized and modularized product platform architectures," *Research in Engineering Design*, vol. 13, no. 3, pp. 213–235, 2002.

[2] Z. Kong and D. Ceglarek, "Fixture workspace synthesis for reconfigurable assembly using procrustes-based pairwise configuration optimization," *Journal of Manufacturing Systems*, vol. 25, no. 1, pp. 25–38, 2006.

[3] A. M. Choubey, P. Prakash, F. T. S. Chan, and M. K. Tiwari, "Solving a fixture configuration design problem using genetic algorithm with learning automata approach," *International Journal of Production Research*, vol. 43, no. 22, pp. 4721–4743, 2005.

[4] K. Bastani, Z. Kong, W. Huang, X. Huo, and Y. Zhou, "Fault diagnosis using an enhanced Relevance Vector Machine (RVM) for partially diagnosable multistation assembly processes," *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 1, pp. 124–136, 2013.

[5] J. Camelio, S. J. Hu, and D. Ceglarek, "Modeling variation propagation of multi-station assembly systems with compliant parts," *Journal of Mechanical Design, Transactions of the ASME*, vol. 125, no. 4, pp. 673–681, 2003.

[6] Z. Kong, D. Ceglarek, and W. Huang, "Multiple fault diagnosis method in multistation assembly processes using orthogonal diagonalization analysis," *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, vol. 130, pp. 1–10, 2008.

[7] Y. Liu, S. Jin, Z. Lin, C. Zheng, and K. Yu, "Optimal sensor placement for fixture fault diagnosis using Bayesian network," *Assembly Automation*, vol. 31, no. 2, pp. 176–181, 2011.

[8] Z. Kong and D. Ceglarek, "Rapid deployment of reconfigurable assembly fixtures using workspace synthesis and visibility analysis," *CIRP Annals*, vol. 52, no. 1, pp. 13–16, 2003.

[9] A. Millar and H. Kihlman, "Reconfigurable flexible tooling for aerospace wing assembly," in *Proceedings of the SAE AeroTech Congress & Exhibition*, Seattle, Wash, USA, November 2009.

[10] D. Tang, W. Eversheim, G. Schuh, and K.-S. Chin, "CyberStamping: a web-based environment for cooperative and integrated stamping product development," *International Journal of Computer Integrated Manufacturing*, vol. 17, no. 6, pp. 504–519, 2004.

[11] D. Tang, "An agent-based collaborative design system to facilitate active die-maker involvement in stamping part design," *Computers in Industry*, vol. 54, no. 3, pp. 253–271, 2004.

[12] W. Eversheim, M. Westekemper, and F. Spennemann, "Cooperation with tool and fixture supplier," in *Strategic Production Networks*, L. Zheng and F. Posseloleken, Eds., pp. 176–187, Springer, Berlin, Germany, 2002.

[13] P. C. Stadzisz and J.-M. Henrioud, "An integrated approach for the design of multi-product assembly systems," *Computers in Industry*, vol. 36, no. 1-2, pp. 21–29, 1998.

[14] E. Travaini, P. Pedrazzoli, R. Rinaldi, and C. R. Boër, "Methodological approach and reconfiguration tool for assembly systems," *CIRP Annals*, vol. 51, no. 1, pp. 9–13, 2002.

[15] Y. Rong and Y. Bai, "Machining accuracy analysis for computer-aided fixture design verification," *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, vol. 118, no. 3, pp. 289–300, 1996.

[16] N. Kaya, "Machining fixture locating and clamping position optimization using genetic algorithms," *Computers in Industry*, vol. 57, no. 2, pp. 112–120, 2006.

[17] D. Tang and B. Gao, "Feature-based metal stamping part and process design—part I: stampability evaluation," *International Journal of Production Research*, vol. 45, no. 12, pp. 2673–2695, 2007.

[18] G. C. Vosniakos and T. Giannakakis, "A knowledge-based manufacturing advisor for pressworked sheet metal parts," *Journal of Intelligence Manufacturing*, 2012.

[19] S. Kumar, R. Singh, and G. S. Sekhon, "CCKBS: a component check knowledge-based system for assessing manufacturability of sheet metal parts," *Journal of Materials Processing Technology*, vol. 172, no. 1, pp. 64–69, 2006.

[20] W. Hu, L. G. Yao, and Z. Z. Hua, "Optimization of sheet metal forming processes by adaptive response surface based on intelligent sampling method," *Journal of Materials Processing Technology*, vol. 197, no. 1–3, pp. 77–88, 2008.

[21] M. Firat, "Computer aided analysis and design of sheet metal forming processes—part 3: stamping die-face design," *Materials and Design*, vol. 28, no. 4, pp. 1311–1320, 2007.

[22] J.-M. Huang, Y.-T. Jou, L.-C. Zhang, S.-T. Wang, and C.-X. Huang, "A web-based model for developing: a mold base design system," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8356–8367, 2009.

[23] J. Li and D. Su, "Support modules and system structure of web-enabled collaborative environment for design and manufacture," *International Journal of Production Research*, vol. 46, no. 9, pp. 2397–2412, 2008.

[24] Y. Yujun, Z. Jinsong, W. Li, and C. Liping, "Internet-based collaborative product development chain for networked product development," *International Journal of Advanced Manufacturing Technology*, vol. 28, no. 7-8, pp. 845–853, 2006.

[25] G. Q. Huang, J. B. Zhao, and K. L. Mak, "Early supplier involvement in new product development with WeBid: a case study on video conferencing system," *Journal of Computing and Information Science in Engineering*, vol. 3, no. 4, pp. 363–365, 2003.

[26] G. Q. Huang and K. L. Mak, "Brokering the customer-supplier partnership in product design and realization over the World Wide Web," *IIE Transactions*, vol. 35, no. 4, pp. 369–378, 2003.

[27] P. Gupta and R. Nagi, *Optimal partner selection for virtual enterprises in agile manufacturing [thesis]*, Department of Industrial Engineering, SUNY at Buffalo, Buffalo, NY, USA, 1995.

[28] D. Y. Sha and Z. H. Che, "Supply chain network design: partner selection and production/distribution planning using a systematic model," *Journal of the Operational Research Society*, vol. 57, no. 1, pp. 52–62, 2006.

[29] Z. Lin and L. Wang, "Agile partner selection based on meta-heuristic algorithms," in *Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE '08)*, pp. 402–407, October 2008.

[30] H. Jeon, C. Petrie, and M. R. Cutkosky, "JATLite: a Java agent infrastructure with message routing," *IEEE Internet Computing*, vol. 4, no. 2, pp. 87–96, 2000.

[31] L. N. de Castro and J. I. Timmis, *Artificial Immune Systems: A New Computational Intelligence Approach*, Springer, London, UK, 2002.

[32] L. N. De Castro and F. J. Von Zuben, "Learning and optimization using the clonal selection principle," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 3, pp. 239–251, 2002.

[33] C.-Y. Tsai, C.-C. Lo, and A. C. Chang, "Using fuzzy QFD to enhance manufacturing strategic planning," *Journal of the Chinese Institute of Industrial Engineers*, vol. 20, no. 1, pp. 33–41, 2003.

[34] J. Lee, S. J. Hu, and A. C. Ward, "Workspace synthesis for flexible fixturing of stampings," *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, vol. 121, no. 3, pp. 478–484, 1999.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 958747, 9 pages
http://dx.doi.org/10.1155/2013/958747

Research Article

Analysis of Novel Variable Reluctance Resolver with Asymmetric Teeth on the Stator

Chengjun Liu, Ming QI, and Meng Zhao

Department of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 150001, China

Correspondence should be addressed to Chengjun Liu; liucj@hit.edu.cn

Received 15 February 2013; Accepted 22 May 2013

Academic Editor: Dan Zhang

The variable reluctance (VR) resolver is usually used to measure the shaft angle of motor. A novel VR resolver with asymmetric teeth on the stator is proposed to achieve the minimal number of active teeth and eliminate the amplitude imbalance of the output voltages of SIN and COS coils by bringing the fifth compensating tooth. The structure and the total output voltages of signal windings are explicated theoretically in detail. The topology of such a machine just requires 2D finite element analysis (FEA) to accurately predict the machine performance. Results of FEA and principle of VR resolver with asymmetric teeth are in good agreement, and several special relevant problems are studied. Finally, experimental results validate the theoretical analysis and FEA of the proposed VR resolver.

1. Introduction

When vector control, direct torque control, and other control strategies are applied in AC governor system with high dynamic performance, the rotor angular position is required necessarily [1–4]. As electromagnetic rotor position sensor of motor, variable reluctance (VR) resolver, which has simple structure, small axial dimension, high reliability, high amplitude of output voltage, low cost, and relatively high precision, is widely used to realize commutation and control algorithms in numerous applications especially in electric vehicle [5–10].

The conventional VR resolver adopts the sophisticated sinusoidal distributed windings placed on the stator and has slots in both stator and rotor to output two quadrature signals [11]. However, the turns of reference winding and signal windings should be an integral number, and thus the angle measurement accuracy will be reduced [12, 13]. Therefore, the novel VR resolver has concentrated windings used for SIN and COS coils and rotor with several salient poles to induce voltages containing the information of the rotor angle [14–16]. Several researches also focus on the resolver models and new-type resolver [17, 18].

Usually, for a conventional 4 pairs of pole VR resolver, the number of teeth on stator is selected as at least 8 or the integer multiple of 8. Unfortunately, when VR resolvers are applied in a small and narrow space; for example, when the stator diameter is less than 30 mm, it is of great difficulty to insert coils because thin copper wires are too fragile and the space inside the resolver for inserting coils manually is too limited to accomplish accurately and conveniently. Thus, a novel VR resolver with asymmetric teeth on the stator, which has a minimal number of active teeth and the simplest structure, is proposed to reduce the size and cost of the device while simplifying its construction without loss of accuracy.

This paper is organized as follows. Firstly, the construction and work principle of novel VR resolver with asymmetric teeth on the stator are introduced in Section 2. The expressions of air-gap permeance and the output voltages of signal windings are deduced. Section 3 shows results of FEA to validate the analysis of novel VR resolver with asymmetric teeth. Related problems including the difference between SIN and COS compensating coils and arbitrariness of the location chosen for the compensating tooth are emphasized. The prototype and test bench are developed based on the analysis and computation results above; the performance of VR resolver with asymmetric teeth is investigated in Section 4, including the output voltages of signal windings and zero-error. Conclusions are summarized in Section 5.

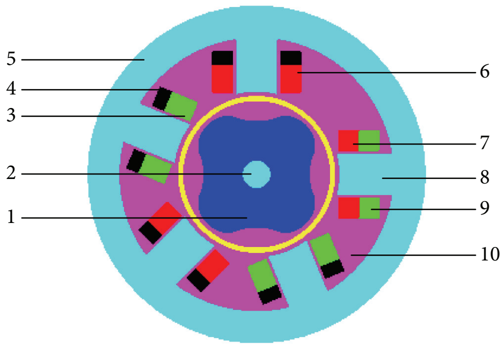


Figure 1: Structure of novel VR resolver with asymmetric teeth on the stator: 1-rotor, 2-shaft, 3-SIN coil, 4-reference winding, 5-stator, 6-COS coil, 7-COS compensating coil, 8-compensating tooth, 9-SIN compensating coil, 10-air.

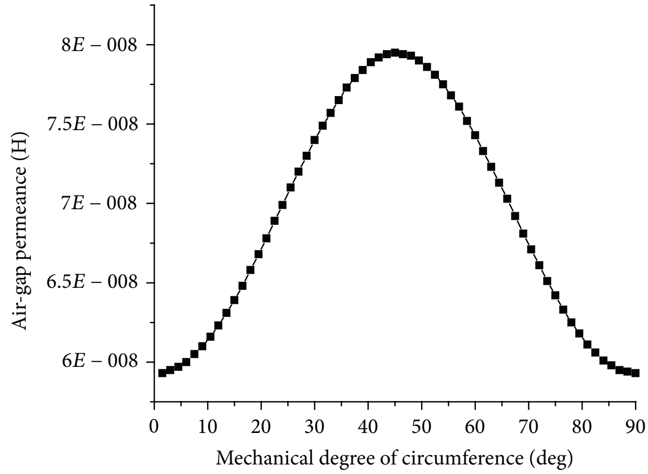


Figure 2: Waveform of air-gap permeance beneath the first tooth.

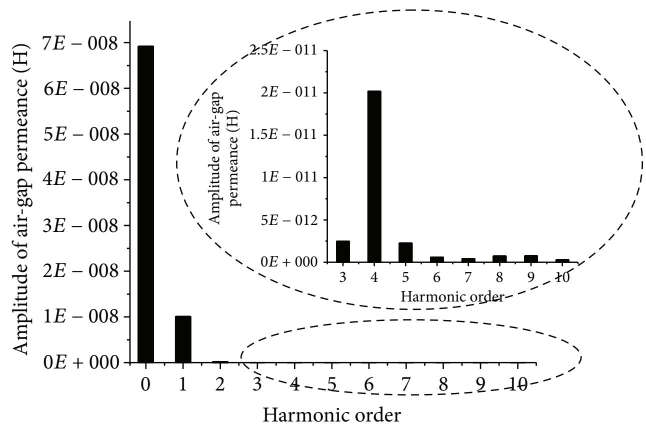


Figure 3: Harmonic components of air-gap permeance.

2. Principle of Novel VR Resolver with Asymmetric Teeth on the Stator

2.1. Structure of VR Resolver with Asymmetric Teeth. The novel variable reluctance resolver, which is completely different from conventional multipole VR resolver, has reference winding and signal windings placed on the stator and several salient poles on the rotor. Thus, due to salient pole effects, with the rotation of the rotor, the inductance of signal windings varies, and the voltages of SIN and COS coils, which contain the information of the rotor angle, can be induced. When the rotor rotates one revolution, the period number of the output signals will be the same as the rotor salient pole number. Normally the higher the rotor pole number is, the more position accuracy the VR resolver can achieve. The stator pole number z should be the times of four to form two symmetrical phases. Moreover, the reference coils are connected in series to form z poles. The polarity of coils will be positive/negative alternatively. In order to make the teeth number be the minimum, four active teeth with reference windings are put on the stator. Another fifth asymmetric tooth, nonactive, is proposed to eliminate the average component from the induced voltage.

The SIN signal should be always in quadrature with COS signal, that is, the SIN winding and the COS winding must be mechanically displaced 90 electrical degrees from each other in the stator. And, the principle of slot-pole combination of VR resolver should be followed strictly. As a result, the number of teeth on the stator is selected as five; the rotor pole number p is chosen as four and the angle-interval of the neighboring tooth α on the stator should be 67.5 mechanical degrees. The structure of novel VR resolver with asymmetric teeth on the stator is shown as Figure 1.

It can be seen that the reference windings and the signal windings are placed together in the stator, and the polarity of reference windings and SIN and COS coils are positive/negative alternatively. GCAACGGATCATCATGTAA The two concentrated coils, called compensating coils, are placed on the compensating tooth and, respectively, connected with SIN and COS coils in series. The stator is not complex and can be manufactured by simply using normal silicon steel and concentrated coils. To form the salient poles more precisely, the rotor may be a little more complicated to be fabricated by wire-electrode cutting. Thus, the novel VR resolver is named as VR resolver with asymmetric teeth for the reason that the angle-interval between the first four neighboring slots is 67.5 mechanical degrees and the fifth is 90 mechanical degrees. The shape of the compensating tooth is just like the other four teeth, and the main difference is that there is no reference winding on it.

2.2. Analysis of the Air-Gap Permeance of VR Resolver with Asymmetric Teeth. The air-gap permeance waveform determines the output voltage waveforms of SIN and COS coils directly. Invariably, the magnetic flux closes along the path where the permeance is the maximum. The main magnetic circuit contains stator yoke, stator teeth, rotor, and air-gap. So, analysis of the permeance under four active teeth should be taken as the principal thing.

According to the general analysis method for a VR resolver, the higher harmonics can be neglected [14]; thus, the

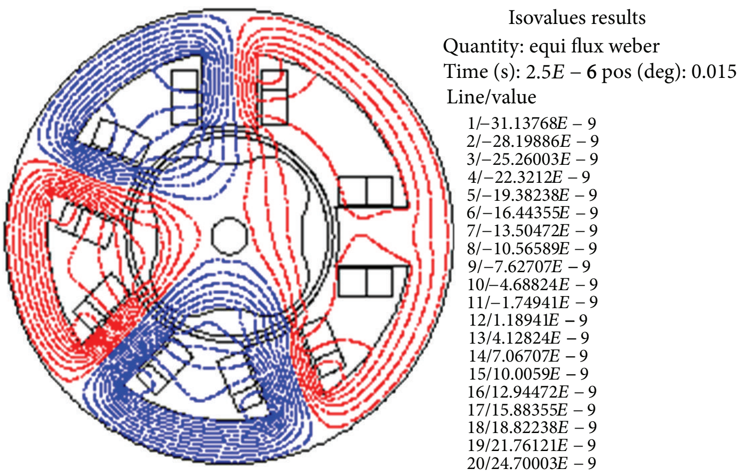


Figure 4: Magnetic flux distribution in the VR resolver with asymmetric teeth on the stator.

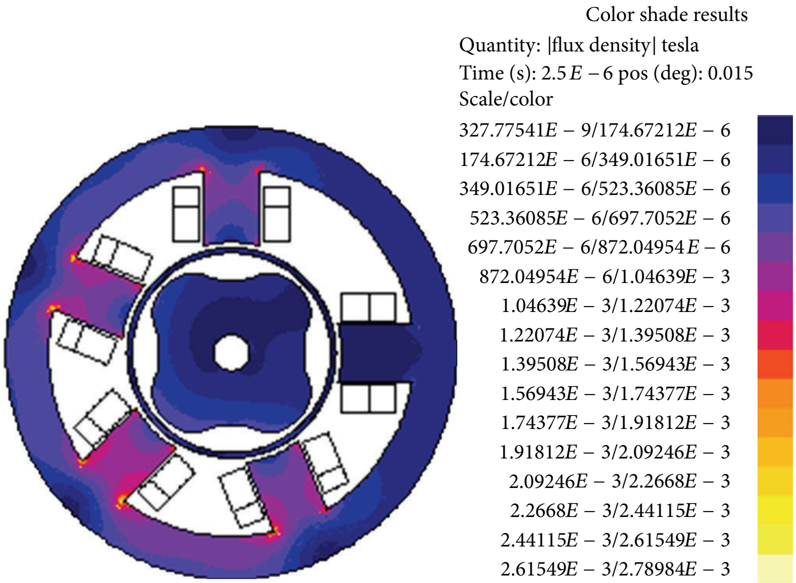


Figure 5: Magnetic flux density in the VR resolver with four teeth.

permeance versus rotor-position angle θ can be expressed as follows:

$$\Lambda_{Zk} = \Lambda_0 + \Lambda_1 \cos p(\theta + k \cdot \alpha), \tag{1}$$

where Λ_0 is the average permeance, Λ_1 is the amplitude of the fundamental component, θ is the rotating angle, p is the salient pole pair number of the rotor, α is the mechanical degree of the neighboring tooth, and k is the number of the stator teeth. For example, the number of the teeth at the top is 1, and then other teeth are numbered by a counterclockwise rotation.

Thus, the sum of the air-gap permeance can be expressed as:

$$\sum_{k=1}^{4} \Lambda_{Zk} = 4\Lambda_0 + \sum_{k=0}^{3} \Lambda_1 \cos p(\theta + k \cdot \alpha) = 4\Lambda_0. \tag{2}$$

Therefore, the sum of the air-gap permeance remains unchanged when the rotor rotates. Then, the reactance of the reference winding is a constant, and the amplitude of the induced voltage of reference winding which is excited by a constant voltage power supply is invariable so that no error will be brought in.

On the condition that reference winding is excited by a constant voltage power supply, the following equation should be followed when calculating the sum of the magneto motive force:

$$\sum_{k=1}^{4} F_{zk} = \frac{\sum_{k=1}^{4} \Phi_{zk}}{\sum_{k=1}^{4} \Lambda_{zk}} = \text{const}, \tag{3}$$

where $\sum_{k=1}^{4} \Phi_{zk}$ is the sum of the air-gap magnetic flux.

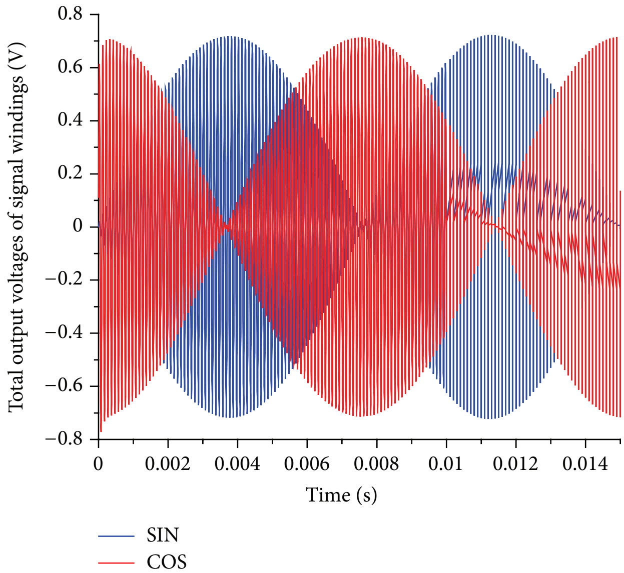


FIGURE 6: Total output voltages of signal windings.

Then, the air-gap magnetic flux of each tooth can be expressed as:

$$\Phi_{zk} = F_{zk} \cdot \Lambda_{zk} = \Phi_0 + \Phi_1 \cos p\,(\theta + k \cdot \alpha)\,, \quad (4)$$

where Φ_0 is the average magnetic flux and Φ_1 is the amplitude of the fundamental component.

2.3. Analysis of the Output Voltages of SIN and COS Coils. To simplify the problem, it is assumed that

(1) the permeance of stator and rotor iron are infinite;

(2) the iron losses are omitted.

The output voltage of SIN coil will turn out to be

$$E_s = \sum_{k=2n}^{4} 4.44 f N_s \cos k\pi \Phi_{zk} \quad (n = 1, 2)\,, \quad (5)$$

where N_s is the turns of each SIN coil, which is the concentrated coil like the reference winding, f is the frequency of the reference signal, and $\cos k\pi$ represents the N/S polarity of reference coils.

Just like the air-gap permeance and the magnetic flux, the output voltage of SIN coil contains average component and fundamental component. The average component, which only varies with time and without the rotation of the rotor, is expressed as:

$$E'_{s0} = \sum_{k=2n}^{4} 4.44 f N_s \cos k\pi \Phi_0 = 0. \quad (6)$$

Nevertheless, the average value of signal windings cannot be neglected because of the following reasons.

(1) The reference winding and signal windings are put together on the stator tooth and magnetic coupling exists inevitably.

(2) There will be direct coupling between the windings due to the stray capacitance effect.

(3) The slot leakage flux will induce voltages irrelevant to the position signal.

(4) The magnetic circuit of VR resolver with asymmetric teeth is typically asymmetrical.

(5) The compensating effect of the signal windings in series will be little in that there exists only four active teeth.

Therefore, the average component of SIN coil should be revised as:

$$E_{s0} = Q_s f N_s \Phi_0, \quad (7)$$

where Q_s is the voltage coefficient of SIN coil and can be calculated by FEA.

The fundamental component will turn out to be

$$\begin{aligned} E_{s1} &= \sum_{k=2n}^{4} 4.44 f N_s \Phi_1 \cos k\pi \cos\left[p\theta + (k-1)\,\alpha\right] \\ &= 8.88 f N_S \Phi_1 \sin p\theta. \end{aligned} \quad (8)$$

Thus, the output voltage of SIN coil could be expressed in a more distinct way as:

$$E_s = E_{s0} + E_{s1} = Q_s f N_S \Phi_0 + 8.88 f N_S \Phi_1 \sin p\theta. \quad (9)$$

Similarly, the output voltage of the COS coil will turn out to be

$$E_c = E_{c0} + E_{c1} = Q_c f N_c \Phi_0 + 8.88 f N_c \Phi_1 \cos p\theta, \quad (10)$$

where N_c is the turns of each concentrated COS coil and Q_c is the voltage coefficient of COS coil and can also be calculated by FEA.

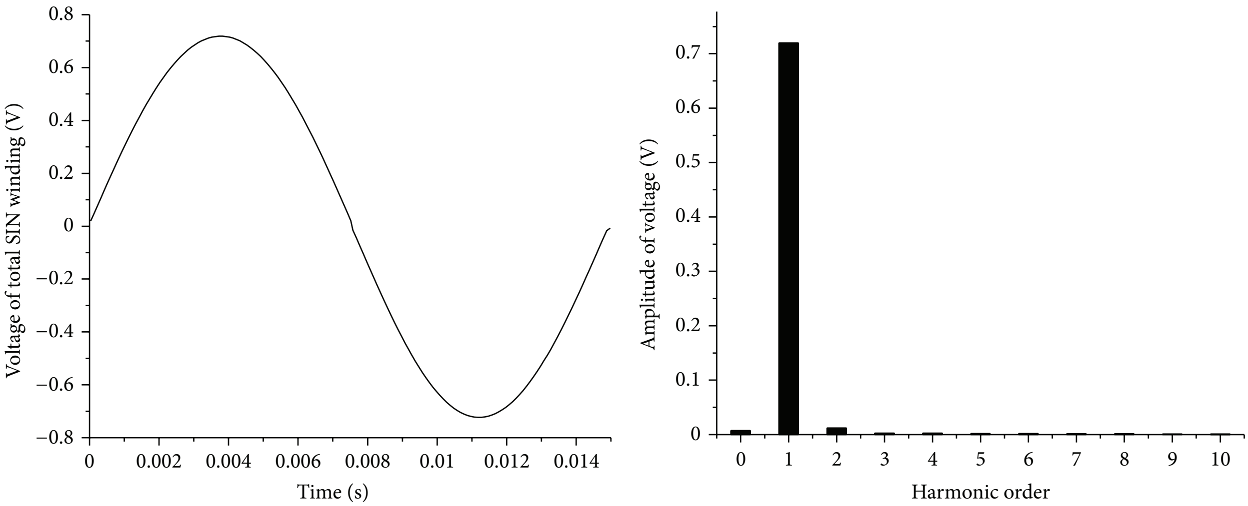
It can be clearly seen from (9) and (10) that the output voltages of VR resolver with asymmetric teeth contain average component and fundamental component. The average component will definitely cause function error and amplitude imbalance when the resolver measures the angle of the rotating shaft. Therefore, the output voltages of VR resolver with asymmetric teeth must be in series with compensating coils to eliminate the average value and obtain higher precision.

2.4. Analysis of the Total Output Voltages of Signal Windings. To simplify the issue, it is assumed that the magnetic flux in the air-gap under the compensating tooth is zero because there is no reference winding placed on it.

Then, based on the principle of continuity of magnetic flux, at any time, the algebraic sum of the magnetic flux in the compensating tooth and the other four active teeth should be zero.

Thus, the magnetic flux in the compensating tooth will turn out to be

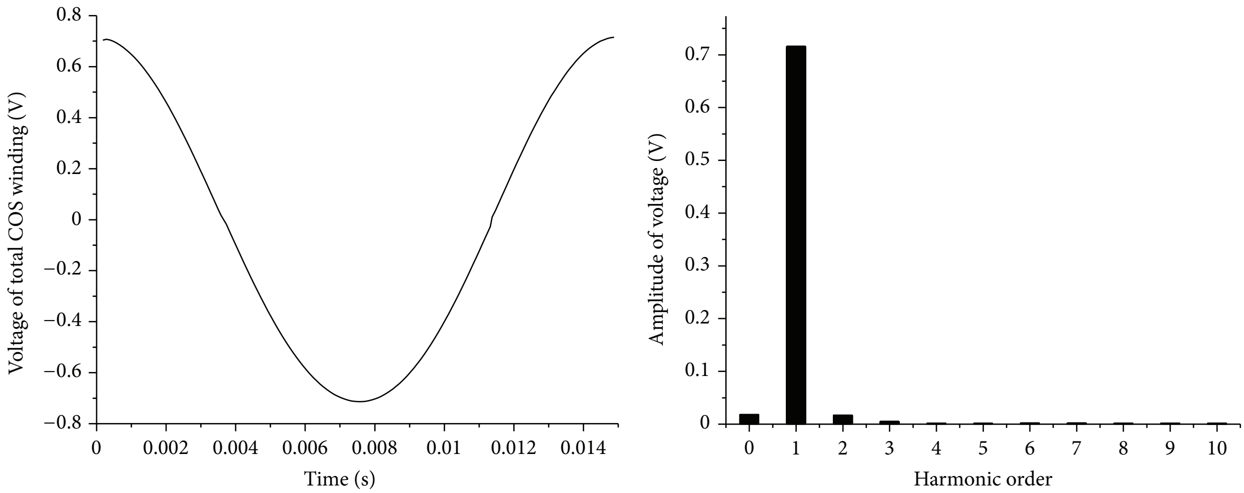
$$\Phi_c = -\sum_{k=1}^{4} \Phi_{zk}. \quad (11)$$



(a) Envelope curve

(b) Harmonic components

Figure 7: Voltage of total SIN winding.



(a) Envelope curve

(b) Harmonic components

Figure 8: Voltage of total COS winding.

The average component of the output voltage of the SIN compensating coil can be expressed as:

$$E_{cs0} = -4.44 f N_{cs} \left(\sum_{k=1}^{4} \Phi_0 \right) = -17.76 f N_{cs} \Phi_0, \qquad (12)$$

where N_{cs} is the turns of the SIN compensating coil.

The function of the compensating coils is to eliminate the average value from the induced voltages of SIN and COS coils, and the turns of compensating coils can only be calculated by FEA accurately.

The fundamental component of output voltage of the SIN compensating coil will turn out to be

$$E_{cs1} = -4.44 f N_0 \sum_{k=1}^{4} \Phi_1 \cos [4\theta + (k-1)\alpha] = 0. \qquad (13)$$

Thus, by connecting the SIN coil with the SIN compensating coil in series, the total output voltage will be

$$E_s' = E_s + E_{cs} = 8.88 f N_s \Phi_1 \sin p\theta. \qquad (14)$$

Similarly, the total output voltages of the COS coil and the COS compensating coil will be

$$E_c' = E_c + E_{cc} = 8.88 f N_c \Phi_1 \cos p\theta. \qquad (15)$$

Obviously, after the compensating tooth and compensating coils are proposed, the total output voltages of the signal windings will only contain fundamental component and vary with the rotation of the rotor strictly, which can naturally obtain a higher precision to a great extent. This is the principle of novel VR resolver with asymmetric teeth on the stator.

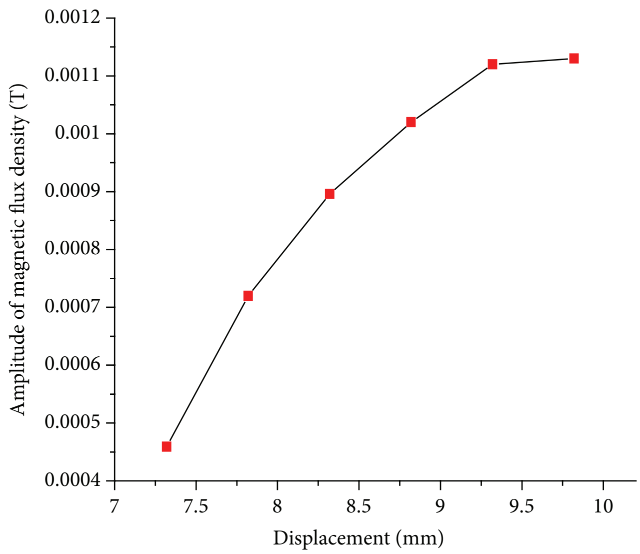


Figure 9: Magnetic flux density in the compensating tooth versus radial displacement along the symmetric line of compensating tooth.

Table 1: Specification of VR resolver with asymmetric teeth.

Items	Units	Specification
Exciting reference voltage	Vrms	7
Exciting frequency	kHz	10
Salient pole pair		4
Minimum air-gap height	mm	1
Axial length	mm	15
Stator outer diameter	mm	28
Turns of each reference coil		20
Turns of each signal coil		20
Turns of the SIN compensating coil		17
Turns of the COS compensating coil		42
Voltage ratio		0.072
Rotating speed	rpm	1000
Calculation time instant	μs	2.5

3. Finite Element Analysis

FEA is used to simulate the performance of novel VR resolver with asymmetric teeth on the stator. The calculation time instant is related to exciting frequency and number of samples. The specification of novel VR resolver with asymmetric teeth is shown in Table 1.

3.1. Calculation of Air-Gap Permeance. The output voltages of signal windings depend on the air-gap permeance waveform. So, the calculation of air-gap permeance should be in the first place. Figure 2 shows the waveform of air-gap permeance beneath the first tooth. The harmonic components are shown in Figure 3, and the total harmonic distortion (THD) is 1.84%. Results of the calculation of permeance validate the analysis of the air-gap permeance of VR resolver with asymmetric teeth on the stator.

Table 2: Calculated results of the total output voltages of signal windings.

Items	Units	Specification
Maximum voltage value of total SIN winding	V	0.7186
Minimum voltage value of total SIN winding	V	−0.7232
Average value of total SIN winding	V	0.0072
Maximum voltage value of total COS winding	V	0.7154
Minimum voltage value of total COS winding	V	−0.7136
Average value of total COS winding	V	0.0182
Electrical error	Degree	0.6

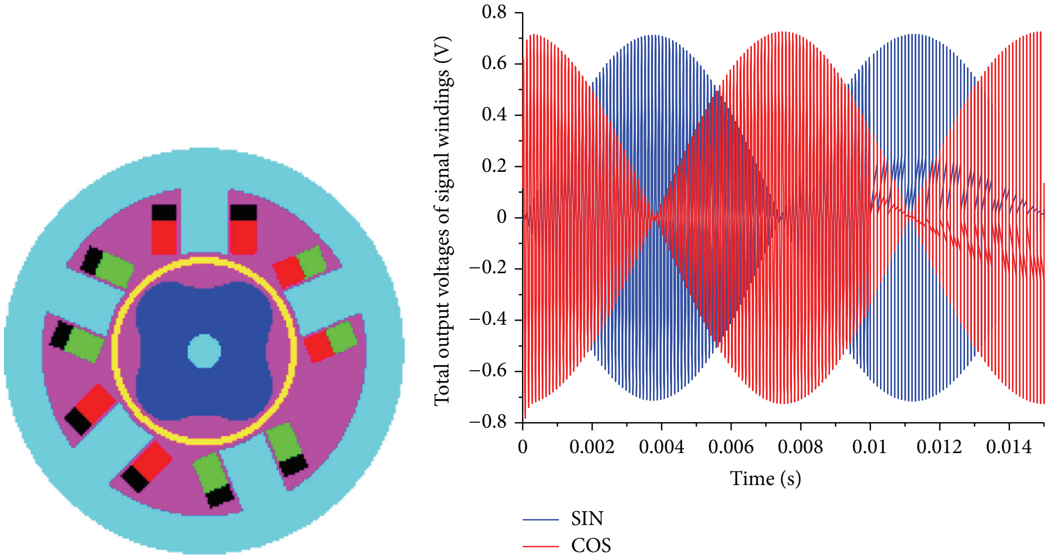
3.2. Calculation of the Total Output Voltages of Signal Windings. Figure 4 presents the magnetic flux distribution in the VR resolver with asymmetric teeth at 2.5 μs. It can be clearly seen from Figure 4 that the main magnetic circuit is asymmetrical, which will induce null voltages irrelevant to the position signal in the output windings to some extent. There exists a little amount of the slot leakage flux compared with the main flux and no magnetic flux under the compensating tooth. The magnetic flux density is shown in Figure 5, and the flux density is relatively low, which will not cause saturation in the stator and rotor. The magnetic flux density in the compensating tooth is lower than the other four teeth as a result of no reference winding on it.

The total output voltages of signal windings are almost equal between the positive and negative half of voltage cycle, as presented in Figure 6. After bringing the compensating tooth, the average value and amplitude imbalance can be reduced significantly and the precision will be improved effectively. Figure 7 shows the envelope curve and harmonic components of total voltage of SIN winding. Figure 8 displays the envelope curve and harmonic components of total voltage of COS winding.

Calculated results of the total output voltages of signal windings are shown in Table 2. Total SIN winding here includes SIN coil and SIN compensating coil and so does total COS winding.

It can be seen from Table 2 that the average value of voltage of total SIN winding and that of total COS winding are different because the turns of SIN and COS compensating coils are not equal. It should be noted that the average components can never be zero because the turns of the compensating coils should be an integral number. The explanation for the reason why SIN and COS compensating coils are not equal will be demonstrated in Section 3.3.

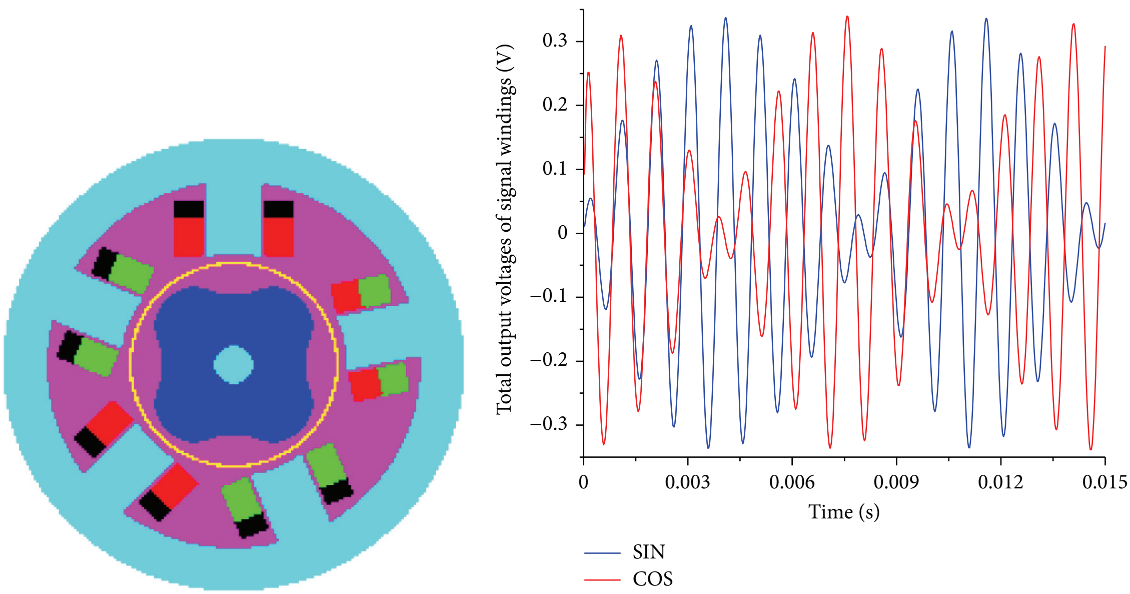
3.3. Explanation for the Difference between SIN and COS Compensating Coils. The turns of SIN compensating coil equal 17 and that of COS compensating coil are 42, as shown in Table 1. Figure 9 presents the magnetic flux density in the compensating tooth versus radial displacement along the symmetric line of compensating tooth. It is supposed that the magnetic flux in the air gap under the compensating tooth



(a) Structure

(b) Output voltages of signal windings

FIGURE 10: VR resolver which the angle-interval between the fourth tooth and compensating tooth is 90 mechanical degrees.



(a) Structure

(b) Output voltages of signal windings

FIGURE 11: VR resolver which the angle-interval between the fourth tooth and compensating tooth is 78.75 mechanical degrees.

should be zero. The magnetic flux density in the compensating tooth varies with displacement along the symmetric line of compensating tooth. The amplitude of magnetic flux density will increase when the radial displacement is closer to the stator yoke. The turns of COS compensating coil is more than that of SIN compensating coil because COS compensating coil is farther away from the stator yoke.

3.4. Arbitrariness of the Location Chosen for the Compensating Tooth. Based on the principle of continuity of magnetic flux, expression (11) shows that the magnetic flux in the compensating tooth is irrelevant to the location of the fifth tooth. Therefore, the location chosen for the compensating tooth is arbitrary. Two typical examples are researched to validate the analysis above. Figure 10 shows the construction and output voltages of signal windings of one example which the angle-interval between the fourth tooth and compensating tooth is 90 mechanical degrees. The structure and output voltages of signal windings of another example which the angle-interval between the fourth tooth and fifth tooth is 78.75 mechanical degrees are shown in Figure 11.

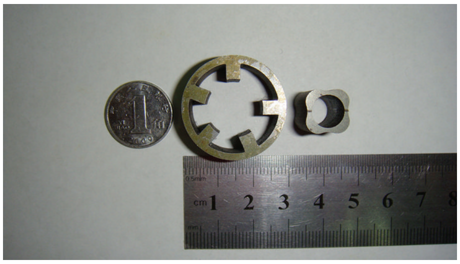


Figure 12: Prototype of novel VR resolver with asymmetric teeth on the stator.

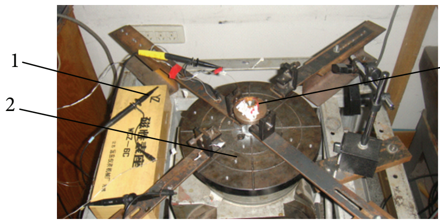


Figure 13: Experiment setup: 1-oscilloscope probe, 2-turntable, 3-prototype.

Both two calculation examples are able to output two quadrature and amplitude-equal signals. The only difference between the three resolvers is that the turns of SIN and COS compensating coil are not invariable because of the distinctions of slot leakage flux distribution and the magnetic flux density in the compensating tooth.

4. Experiment and Error Analysis

4.1. Prototype. Based on the analysis and computations above, a prototype has been made as shown in Figure 12, with four active teeth and one nonactive tooth on the stator and four salient poles on the rotor.

4.2. Output Voltages of Signal Windings Measurement. Figure 13 presents the test bench for output voltages of signal windings measurement. The stator was fixed on the support, and the rotor was rotating with the shaft fixed on the turntable, and thus their relative positions can be changed. The output voltages of signal windings are measured by TEK oscilloscope, as shown in Figure 14. The experimental results prove the validities of the analysis and results of FEA above.

4.3. Zero-Error Measurement. Figure 15 shows the zero-error within a mechanical cycle which is measured by high-precision digital grating optical dividing head.

It can be seen that the maximum zero-error of SIN winding is ±0.66° and that of COS winding is ±0.71°. The zero-error changes in a zigzag fashion at different zero positions because noneffective voltages exist in output signals. Compared with electrical error computed by FEA, results of the experiment is larger for the reason that it is difficult to make the stator and rotor absolutely concentric in the experiment.

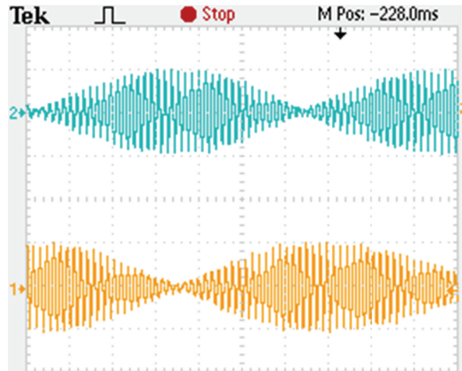


Figure 14: Oscilloscope graph measured: 1-voltage of COS winding, 2-voltage of SIN winding.

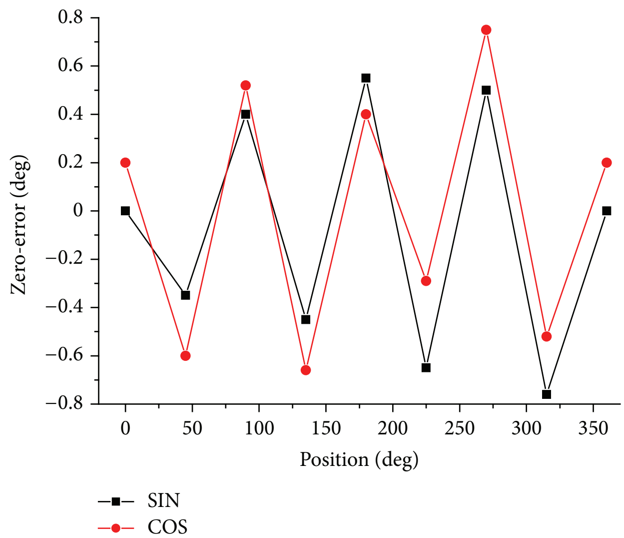


Figure 15: Zero-error versus rotating angle.

5. Conclusions

The following conclusions can be drawn from the analysis of the novel VR resolver with asymmetric teeth on the stator.

(1) The fifth asymmetric tooth, nonactive, called compensating tooth, plays a pivotal role in the new topology of VR resolver. The induced voltages of compensating coils can eliminate the average value from the voltages of SIN and COS coils so that the precision will be improved substantially.

(2) The calculated electric error of the proposed VR resolver can approach ±0.7°. Several relevant special issues caused by special construction are researched in detail. Computations of FEA are consistent with the principle analysis of the proposed VR resolver.

(3) Experimental results show that the novel VR resolver has equivalent precision (±0.5°) compared with the same type of conventional top VR resolvers. However, it is confirmed that the novel VR resolver with asymmetric teeth on the stator is advantaged with

reliability and cost due to the simple manufacturing process and structure of stator. It is more cost effective, so, as position sensor, it is very useful for low cost drives or electric vehicles.

Conflict of Interests

The authors have stated explicitly that there is no conflict of interests in connection with this paper.

Acknowledgments

This work was supported by National Natural Science Foundation of China (51107013), Natural Scientific Research Innovation Foundation of Harbin Institute of Technology (HIT.NSRIF.2011141), and 973 Program under Grant No. 2013CB035605.

References

[1] G. Pellegrino, R. I. Bojoi, and P. Guglielmi, "Unified direct-flux vector control for AC motor drives," *IEEE Transactions on Industry Applications*, vol. 47, no. 5, pp. 2093–2102, 2011.

[2] C. B. Jacobina, E. R. C. da Silva, and A. M. N. Lima, "Vector control strategies for single-phase induction motor drive systems," *IEEE Transactions on Industrial Electronics*, vol. 51, no. 5, pp. 1073–1080, 2004.

[3] S. Kuo-Kai, L. Jun-Kuh, and P. Van-Truong, "Global minimum torque ripple design for direct torque control of induction motor drives," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 9, pp. 3148–3156, 2010.

[4] C. Patel, P. P. Rajeevan, A. Dey, R. Ramchand, K. Gopakumar, and M. P. Kazmierkowski, "Fast direct torque control of an open-end induction motor drive using 12-sided polygonal voltage space vectors," *IEEE Transactions on Power Electronics*, vol. 27, no. 1, pp. 400–410, 2012.

[5] K.-C. Kim, C. S. Jin, and J. Lee, "Magnetic shield design between interior permanent magnet synchronous motor and sensor for hybrid electric vehicle," *IEEE Transactions on Magnetics*, vol. 45, no. 6, pp. 2835–2838, 2009.

[6] K.-C. Kim, S.-J. Hwang, K.-Y. Sung, and Y. Kim, "A study on the fault diagnosis analysis of variable reluctance resolver for electric vehicle," in *Proceedings of the 9th IEEE Sensors Conference (SENSORS '10)*, pp. 290–295, November 2010.

[7] L. Shao, Z. Tang, K. Maki, H. Funato, J. Moore, and G. Saikalis, "Integrated simulation and analysis of resolver sub-system for HEV electric drive," in *Proceedings of the IEEE Vehicle Power and Propulsion Conference (VPPC '08)*, pp. 1–5, September 2008.

[8] A. Bünte and S. Beineke, "High-Performance speed measurement by suppression of systematic resolver and encoder errors," *IEEE Transactions on Industrial Electronics*, vol. 51, no. 1, pp. 49–53, 2004.

[9] M. Konghirun, "A resolver-based vector control drive of permanent magnet synchronous motor on a fixed-point digital signal processor," in *Proceedings of the IEEE Region 10 Conference: Analog and Digital Techniques in Electrical Engineering*, pp. D167–D170, November 2004.

[10] A. Murray, B. Hare, and A. Hirao, "Resolver position sensing system with integrated fault detection for automotive applications," in *Proceedings of the 1st IEEE International Conference on Sensors (IEEE Sensors '02)*, pp. 864–869, June 2002.

[11] D. C. Hanselman, R. E. Thibodeau, and D. J. Smith, "Variable-reluctance resolver design guidelines," in *Proceedings of the 15th Annual Conference of IEEE Industrial Electronics Society (IECON '89)*, pp. 203–208, November 1989.

[12] L. Knox, "Synchro and resolver performance definitions," *IRE Transactions on Component Parts*, vol. 3, no. 3, pp. 88–98.

[13] D. C. Hanselman, "Resolver signal requirements for high accuracy resolver-to-digital conversion," *IEEE Transactions on Industrial Electronics*, vol. 37, no. 6, pp. 556–561, 1990.

[14] L. Sun and Y. Lu, "Rotor-position sensing system based on one type of variable-reluctance resolver," in *Proceedings of the 32nd Annual Conference on IEEE Industrial Electronics (IECON '06)*, pp. 1162–1165, November 2006.

[15] L. Sun, "Analysis and improvement on the structure of variable reluctance resolvers," *IEEE Transactions on Magnetics*, vol. 44, no. 8, pp. 2002–2008, 2008.

[16] L. Sun, J. Shang, and J. Zou, "New absolute rotor-position sensors for inverter-driven motors," in *Proceedings of the Digests of the IEEE International Magnetics Conference (INTERMAG ASIA '05)*, pp. 975–976, Nagoya, Japan, April 2005.

[17] J. Figueiredo, "Resolver models for manufacturing," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 8, pp. 3693–3700, 2011.

[18] F. Kimura, M. Gondo, A. Yamamoto, and T. Higuchi, "Resolver compatible capacitive rotary position sensor," in *Proceedings of the 35th Annual Conference of the IEEE Industrial Electronics Society (IECON '09)*, pp. 1923–1928, November 2009.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 149046, 12 pages
http://dx.doi.org/10.1155/2013/149046

Research Article

Nonlinear Dynamic Analysis and Optimization of Closed-Form Planetary Gear System

Qilin Huang,[1,2] Yong Wang,[1,2] Zhipu Huo,[1,2] and Yudong Xie[1,2]

[1] *School of Mechanical Engineering, Shandong University, Jingshi Road 17923, Jinan, Shandong 25006, China*
[2] *Key Laboratory of High-Efficiency and Clean Mechanical Manufacture, Shandong University, Ministry of Education, Jinan, Shandong 250061, China*

Correspondence should be addressed to Yong Wang; meywang@sdu.edu.cn

Received 3 February 2013; Accepted 10 May 2013

Academic Editor: Zhuming Bi

A nonlinear purely rotational dynamic model of a multistage closed-form planetary gear set formed by two simple planetary stages is proposed in this study. The model includes time-varying mesh stiffness, excitation fluctuation and gear backlash nonlinearities. The nonlinear differential equations of motion are solved numerically using variable step-size Runge-Kutta. In order to obtain function expression of optimization objective, the nonlinear differential equations of motion are solved analytically using harmonic balance method (HBM). Based on the analytical solution of dynamic equations, the optimization mathematical model which aims at minimizing the vibration displacement of the low-speed carrier and the total mass of the gear transmission system is established. The optimization toolbox in MATLAB program is adopted to obtain the optimal solution. A case is studied to demonstrate the effectiveness of the dynamic model and the optimization method. The results show that the dynamic properties of the closed-form planetary gear transmission system have been improved and the total mass of the gear set has been decreased significantly.

1. Introduction

Planetary gear sets have been widely used in engineering including automotive transmissions, aviation transmissions, and crane gearboxes as well as other marine and industrial power transmission systems. Planetary gear trains have many advantages over fixed-center counter-shaft gear systems. The flow of power via multiple-gear meshes increases the power density and helps to reduce the overall size of the transmission train. The ability of multistage planetary sets in providing multiple speed reduction ratios has been the main reason for their extensive use in automatic transmission applications. Closed-form planetary trains are obtained from a number of single-stage differential planetary gear sets and one quasi-planetary stage whose central members are connected according to a given power flow configuration. Input, output, and fixed member assignments are made to certain central members to achieve a given gear ratio.

Because of complexity of structure, most of the earlier published studies on the planetary gear systems were confined to single-stage planetary. In addition, these early models were of linear time-invariant type, so that the eigen solutions and model summation techniques were used to predict the natural modes and the forced response [1–3]. Kahraman [4] employed a purely rotational dynamic model for all possible power flow configurations of complex compound planetary gear sets. He classified the natural model in two categories: asymmetric planet modes and axi-symmetric overall modes. Sun and Hu [5] investigated the frequency response of nonlinear planetary transmission system with multiple clearances using single-term harmonic balance method and focusing only on a single power flow configuration, in which the ring gear was fixed. Al-shyyab and Kahraman [6] developed a rotational single-stage nonlinear dynamic model of a simple planetary gear set and provided a semi-analytical forced response solution using multiterm HBM and showed that these HBM solutions in well agreement with direct numerical integration solution. Also, a recent study by Al-shyyab [7] investigated a compound planetary gear set formed by any number of simple planetary stages, and each

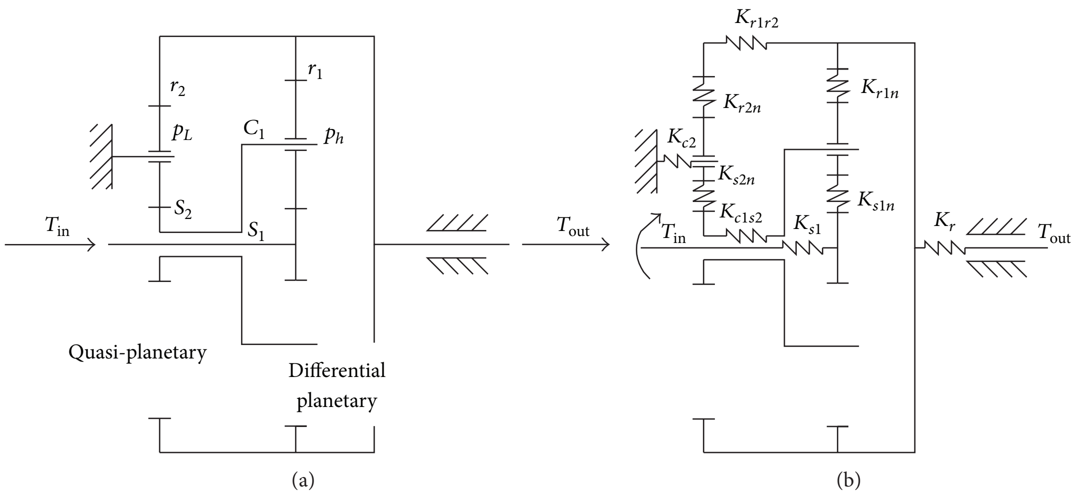


Figure 1: (a) A closed-form planetary gear system; (b) A discrete model of closed-form planetary gear system.

planetary stage has a distinct fundamental mesh frequency and any number of planets spaced in any angular positions using multiterm HBM.

Those studies cited previously are mainly aimed at the modeling of planetary gear trains, the analysis of dynamic response as well as the analysis of parameters stability, and so on. Whereas, the studies on nonlinear dynamic optimization design of planetary transmission system with multiple clearances are still very limited. Zeng Bao [8] and Guan Wei [9] investigated multistage helical gear trains and single-stage planetary gear train taking the dynamic properties as objective functions, respectively. But, in their studies the design variables of optimization were limited to these parameters: the number of gears, the pitch-cycle helical-angle, and the modification coefficients. This study aims at providing numerical solutions and analytical solutions for the dynamic response of a closed-form planetary gear train having three planets spaced equally position angle using Runge-Kutta numerical integration and HBM, respectively. Based on the analytical solution, the optimization mathematical model that focused on minimizing the vibration acceleration of structure and the total mass of the gear transmission system is established. Some key design parameters such as the number of each gear, the module of each stage planetary, the transmission ratio of each stage planetary and the pressure angle are chosen as design varies. This study is available for the designing of closed-form planetary gear sets of both minimum weight and best dynamic characteristic for reference.

2. Dynamic Model of System

2.1. Model and Assumptions. The closed-form planetary gear train consists of a single-stage differential planetary (low-speed stage) and a single-stage quasi-planetary (high-speed stage, carrier is fixed) in this study, as shown in Figure 1. Each stage is comprised of three central elements: the sun gear (s_1, s_2), the ring gear (r_1, r_2), and carrier (c_1, c_2). Each stage planetary has n planet gears. The parameter n representing the number of planet gears is taken as 3 throughout this paper. The planets of each stage are free to rotate with respect to their common carrier. All the gears are mounted on their rigid shafts supported by rolling element bearings. The two rings are connected by torsional linear springs of stiffness k_{r1r2}, as shown in Figure 2. Likewise, the other central elements c_1, c_2 and s_1, s_2 are constrained by torsional linear springs of stiffness k_{c1s2}, k_{c2} and k_{s1}, k_{c1s2}, respectively. In order to establish the mathematical model of the transmission system, a number of simplified assumptions are introduced in the case of speed reduction in the closed-form planetary gear set, as shown in Figure 1.

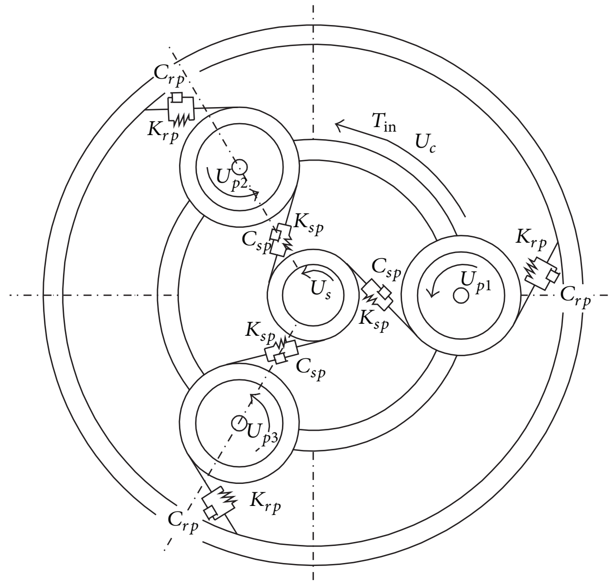


Figure 2: Lumped parameter dynamic model of single-stage planetary gear set.

(1) All of the gears in the set are assumed to be rigid and the flexibilities of each gear teeth at the gear mesh interface are modeled by an equivalent spring having time-varying stiffness acting along the mesh directions. These periodically time-varying mesh stiffnesses are subject to piece-linear functions representing gear backlashes.

(2) Because the bending stiffness of shafts in the set is very large, the deflection of these shafts can be neglected. Thus, the transverse displacements of gears are not considered.

(3) Because the damping mechanisms at the gear meshes and bearings of a planetary gear set are not easy to give a description of mathematical model, viscous gear mesh damping elements are introduced to represent energy dissipation of the transmission system.

2.2. Equivalent Displacements. In order to establish the equations of motion easily, all of torsional angular displacements are unified on the pressure line in terms of equivalent displacements. The equivalent transverse displacements in the mesh line direction caused by rotational displacements are written as follows:

$$u_{cj} = r_{cj}\theta_{cj}, \quad u_{rj} = r_{rj}\theta_{rj}, \quad u_{sj} = r_{sj}\theta_{sj}, \quad u_{pjn} = r_{pj}\theta_{jn}, \quad j = 1,2;\ n = 1,2,3, \tag{1}$$

where θ and r (subscripts cj, rj, sj, jn; $j = h, l$; $n = 1, 2, 3$) are angular displacements and base circle radius of the parts, respectively. While r (subscripts $c1$, $c2$) is the equivalent base circle radius of the carriers defined as follows:

$$\begin{aligned} r_{s1} + r_{ph} &= r_{r1} - r_{ph} = r_{c1}\cos\alpha, \\ r_{s2} + r_{pl} &= r_{r2} - r_{pl} = r_{c2}\cos\alpha. \end{aligned} \tag{2}$$

With the symbol U is used to represent the relative displacements in the direction of pressure line, the relative displacements are obtained according to the meshing relation and the equivalent displacements as follows. The positive direction of the relative displacements is assumed to be the same direction of the compressive deformation. The relative displacements of the parts in the pressure line direction are written as follows:

$$\begin{aligned} U_{s1phi} &= r_{s1}(\theta_{s1} - \theta_{c1}) - r_{ph}(\theta_{hi} + \theta_{c1}) \\ &= r_{s1}\theta_{s1} - r_{c1}\theta_{c1}\cos\alpha - r_{ph}\theta_{hi}, \end{aligned} \tag{3a}$$

$$\begin{aligned} U_{r1phi} &= r_{ph}(\theta_{hi} + \theta_{c1}) - r_{r1}(\theta_{r1} + \theta_{c1}) \\ &= r_{ph}\theta_{hi} - r_{c1}\theta_{c1}\cos\alpha - r_{r1}\theta_{r1}, \end{aligned} \tag{3b}$$

$$U_{c1s2} = r_{c1}(\theta_{c1} - \theta_{s2})\cos\alpha, \tag{3c}$$

$$U_{c2} = r_{c2}\theta_{c2}\cos\alpha, \tag{3d}$$

$$\begin{aligned} U_{s2pli} &= r_{s2}(\theta_{s2} - \theta_{c2}) - r_{pl}(\theta_{li} + \theta_{c2}) \\ &= r_{s2}\theta_{s2} - r_{c2}\theta_{c2}\cos\alpha - r_{pl}\theta_{li}, \end{aligned} \tag{3e}$$

$$\begin{aligned} U_{r2pli} &= r_{pl}(\theta_{li} + \theta_{c2}) - r_{r2}(\theta_{r2} + \theta_{c2}) \\ &= r_{pl}\theta_{li} - r_{c2}\theta_{c2}\cos\alpha - r_{r2}\theta_{r2}, \end{aligned} \tag{3f}$$

$$U_{r1r2} = r_{r1}(\theta_{r1} - \theta_{r2}). \tag{3g}$$

2.3. Equations of Motion. The closed-form planetary gear system consists of four different kinds of gear pairs, the external gear pair, that is, the sun gear/planet gear-i pair (subscripts $s1$, phi and $s2$, pli), and the internal gear pair, that is, the ring gear/planet gear-i pair (subscripts $r1$, phi and $r2$, pli). The mesh of gear j (si or ri, $i = 1, 2$) with a planet pi ($i = h, l$) is represented by a periodically time-varying stiffness element $k_{jpi}(t)$ subjected to a piecewise linear backlash function g that includes a clearance of gap width $2b_{jpi}$. Accordingly, the dynamic model of a closed-form gear set with n planets includes $2n$ clearances. In this closed-form planetary gear system, damper is described by constant viscous damper coefficient c_{jpi}. This is a rather simplified mesh contact model; in reality these contacts are subjected to the hydro-elastic-dynamic regime of lubrication [10]. In this paper, it is supposed that all planets pi ($i = h, l$) and their respective meshes with gear j ($s1$, $s2$ or $r1$, $r2$) are identical so that $k_{jpi}(t)$, b_{jpi}, and c_{jpi} are the same for each jpi mesh, except the phase angles of $k_{jpi}(t)$ which differ according to planet phasing conditions.

Thus, the nonlinear dynamic differential equations of motion of closed-form planetary gear set can be established using the Lagrange principle as follows:

$$\begin{aligned} &M_{c1}\ddot{u}_{c1} + C_{c1s2}\dot{U}_{c1s2} \\ &\quad - \sum_{i=1}^{3}\left(C_{s1n}\dot{U}_{s1phi} + C_{r1n}\dot{U}_{r1phi}\right) + k_{c1s2}U_{c1s2} \\ &\quad - \sum_{i=1}^{3}\left[k_{s1n}U_{s1phi} + k_{r1n}U_{r1phi}\right] = 0, \end{aligned} \tag{4a}$$

$$\begin{aligned} &M_{r1}\ddot{u}_{r1} + C_{r1r2}\dot{U}_{r1r2} \\ &\quad - \sum_{i=1}^{3}C_{r1n}\dot{U}_{r1phi} + k_{r1r2}U_{r1r2} \\ &\quad - \sum_{i=1}^{3}k_{r1n}U_{r1phi} = 0, \end{aligned} \tag{4b}$$

$$M_{s1}\ddot{u}_{s1} + \sum_{i=1}^{3}C_{s1n}\dot{U}_{s1phi} + \sum_{i=1}^{3}k_{s1n}U_{s1phi} = \frac{T_{\text{in}}}{r_{s1}}, \tag{4c}$$

$$\begin{aligned} &M_{ph}\ddot{u}_{hn} - C_{s1n}\dot{U}_{s1phn} + C_{r1n}\dot{U}_{r1phn} \\ &\quad - k_{s1n}U_{s1phn} + k_{r1n}U_{r1phn} = 0 \quad (n = 1, 2, 3), \end{aligned} \tag{4d}$$

$$
\begin{aligned}
& M_{c2}\ddot{u}_{c2}+C_{c2}\dot{U}_{c2} \\
& \quad -\sum_{i=1}^{3}\left(C_{s2n}\dot{U}_{s2pli}+C_{r2n}\dot{U}_{r2pli}\right)+k_{c2}U_{c2} \\
& \quad -\sum_{i=1}^{3}\left[k_{s2n}U_{s2\,pli}+k_{r2n}U_{r2\,pli}\right]=0,
\end{aligned} \tag{4e}
$$

$$
\begin{aligned}
& M_{r2}\ddot{u}_{r2}-C_{r1r2}\dot{U}_{r1r2}\frac{r_{r1}}{r_{r2}} \\
& \quad -\sum_{i=1}^{3}C_{r2n}\dot{U}_{r2pli}-k_{r1r2}U_{r1r2}\frac{r_{r1}}{r_{r2}} \\
& \quad -\sum_{i=1}^{3}k_{r2n}U_{r2\,pli}=\frac{T_{\text{out}}}{r_{r2}},
\end{aligned} \tag{4f}
$$

$$
\begin{aligned}
& M_{s2}\ddot{u}_{s2}-C_{c1s2}\dot{U}_{c1s2}\frac{r_{c1}\cos\alpha}{r_{s2}} \\
& \quad +\sum_{i=1}^{3}C_{s2n}\dot{U}_{s2pli}-k_{c1s2}U_{c1s2}\frac{r_{c1}\cos\alpha}{r_{s2}} \\
& \quad +\sum_{i=1}^{3}k_{s2n}U_{s2\,pli}=0,
\end{aligned} \tag{4g}
$$

$$
\begin{aligned}
& M_{pl}\ddot{u}_{ln}-C_{s2n}\dot{U}_{s2pln}+C_{r2n}\dot{U}_{r2pln} \\
& \quad -k_{s2n}U_{s2\,pln}+k_{r2n}U_{r2\,pln}=0\quad(n=1,2,3),
\end{aligned} \tag{4h}
$$

where, $I_{c1}=I'_{c1}+3m_{ph}r_{c1}^2$ is the equivalent mass moment of inertia of the carrier including planet gears in high-speed stage; I'_{c1} is the inertia of the carrier in high-speed stage; m_{ph} is the actual mass of planet-gear in high-speed stage; and r_{c1} is the distribution circle radius of the planet gears. $I_{c1}=M_{c1}r_{bc1}^2$, M_{c1} is the equivalent mass of the carrier in high-speed stage with respect to the its equivalent base circle radius; r_{bc1} is equivalent base circle radius of the carrier in high-speed stage; $r_{bc1}=r_{c1}\cos\alpha$; $I_{r1}=M_{r1}r_{r1}^2$, and M_{r1} is the equivalent mass of the ring in low-speed stage with respect to its base circle radius; $I_{s1}=M_{s1}r_{s1}^2$, and M_{s1} is the equivalent mass of the sun gear in low-speed stage with respect to its base circle radius; $I_{phn}=M_{ph}r_{ph}^2$, and M_{ph} is the equivalent mass of the planet-gear in high-speed stage with respect to its equivalent base circle radius; $I_{c2}=I'_{c2}+3m_{pl}r_{c2}^2$, and I'_{c2} is the moment of inertia of the carrier in low-speed stage.

Dynamic model of motion of transmission has considerable difficulties in its solution procedure as follows. (1) As a semidefinite system, its first-order natural frequency is zero corresponding to a rigid body motion. (2) The piecewise-linear function g represents the gear backlash, and the number of variables is even different according to the external and internal gear pairs. (3) As both linear and nonlinear restoring forces exist in the equations, it is not possible to write out the governing equation in matrix form, while a general solution technique applicable to the systems of multiple degrees of freedom must be based on matrix form.

Therefore, (4a)–(4h) are simplified further by introducing a set of new variables:

$$U_{s1phn}=u_{s1}-u_{c1}-u_{hn}, \tag{5a}$$

$$U_{r1phn}=u_{hn}-u_{c1}-u_{r1}, \tag{5b}$$

$$U_{c1s2}=u_{c1}-\frac{r_{c1}\cos\alpha}{r_{s2}}u_{s2}, \tag{5c}$$

$$U_{c2}=u_{c2}, \tag{5d}$$

$$U_{s2phn}=u_{s2}-u_{c2}-u_{ln}, \tag{5e}$$

$$U_{r2phn}=u_{ln}-u_{c2}-u_{r2}, \tag{5f}$$

$$U_{r1r2}=u_{r1}-\frac{r_{r1}}{r_{r2}}u_{r2}. \tag{5g}$$

The new coordinate variables defined previously not only have intuitional physical meaning, but also eliminate the rigid body motion. Furthermore, the piecewise-linear backlash function g can be written as a set of functions with a single variable according to (5a)–(5g).

In addition, nondimensional parameters of (4a)–(4h) can be obtained by a characteristic length b and frequency $\omega_n=\sqrt{k_{s1n}/M_{s1}}$, such that

$$t=\omega_n\bar{t}. \tag{6}$$

Hence, a set of simplified equations of motions of the transmission system is obtained by substituting (5a)–(5g) and (6) into (4a)–(4h):

$$
\begin{aligned}
& \ddot{U}_{s1phn}+\frac{1}{\omega_n^2}\sum_{i=1}^{3}\left[\frac{C_{s1n}\dot{U}_{s1phi}}{M_{s1c1}}+\frac{C_{r1n}\dot{U}_{r1phi}}{M_{c1}}\right] \\
& \quad +\frac{1}{\omega_n^2}\left[\frac{C_{s1n}\dot{U}_{s1phn}-C_{r1n}\dot{U}_{r1phn}}{M_{ph}}\right. \\
& \qquad \left.-\frac{C_{c1s2}\dot{U}_{c1s2}}{M_{c1}}\right] \\
& \quad +\frac{1}{\omega_n^2}\sum_{i=1}^{3}\left[\frac{k_{s1n}g\left(U_{s1\,phi}\right)}{M_{s1c1}}+\frac{k_{r1n}g\left(U_{r1\,phi}\right)}{M_{c1}}\right] \\
& \quad +\frac{1}{\omega_n^2}\left[\frac{k_{s1n}g\left(U_{s1\,phn}\right)-k_{r1n}g\left(U_{r1\,phn}\right)}{M_{ph}}\right. \\
& \qquad \left.-\frac{k_{c1s2}g\left(U_{c1s2}\right)}{M_{c1}}\right] \\
& = \frac{T_{\text{in}}}{r_{s1}M_{s1}b\omega_n^2}\quad(n=1,2,3),
\end{aligned} \tag{7a}
$$

$$\ddot{U}_{r1phn} + \frac{1}{\omega_n^2}\left[\frac{C_{r1n}\dot{U}_{r1phn} - C_{s1n}\dot{U}_{s1phn}}{M_{ph}} - \frac{C_{c1s2}\dot{U}_{c1s2}}{M_{c1}} - \frac{C_{r1r2}\dot{U}_{r1r2}}{M_{r1}}\right] + \frac{1}{\omega_n^2}\sum_{i=1}^{3}\left[\frac{C_{s1n}\dot{U}_{s1phi}}{M_{c1}} + \frac{C_{r1n}\dot{U}_{r1phi}}{M_{c1r1}}\right] + \frac{1}{\omega_n^2}\sum_{i=1}^{3}\left[\frac{k_{s1n}g\left(U_{s1phi}\right)}{M_{c1}} + \frac{k_{r1n}g\left(U_{r1phi}\right)}{M_{c1r1}}\right] + \frac{1}{\omega_n^2}\left[\frac{k_{r1n}g\left(U_{r1phn}\right) - k_{s1n}g\left(U_{s1phn}\right)}{M_{ph}} - \frac{k_{c1s2}g\left(U_{c1s2}\right)}{M_{c1}} - \frac{k_{r1r2}g\left(U_{r1r2}\right)}{M_{r1}}\right] = 0 \quad (n = 1,2,3), \tag{7b}$$

$$\ddot{U}_{c1s2} + \frac{C_{c1s2}}{M_{c1}\omega_n^2}\dot{U}_{c1s2} - \frac{1}{\omega_n^2}\sum_{i=1}^{3}\left[\frac{C_{s1n}\dot{U}_{s1phi} + C_{r1n}\dot{U}_{r1phi}}{M_{c1}} + \frac{C_{s2n}\dot{U}_{s2pli}r_{c1}\cos\alpha}{M_{s2}r_{s2}}\right] + \frac{C_{c1s2}}{M_{s2}\omega_n^2}\left(\frac{r_{c1}\cos\alpha}{r_{s2}}\right)^2\dot{U}_{c1s2} + \frac{k_{c1s2}g\left(U_{c1s2}\right)}{M_{c1}\omega_n^2} + \frac{k_{c1s2}g\left(U_{c1s2}\right)}{M_{s2}\omega_n^2}\left(\frac{r_{c1}\cos\alpha}{r_{s2}}\right)^2 - \frac{1}{\omega_n^2}\sum_{i=1}^{3}\left[\frac{k_{s1n}g\left(U_{s1phi}\right) + k_{r1n}g\left(U_{r1phi}\right)}{M_{c1}} + \frac{k_{s2n}g\left(U_{s2pli}\right)r_{c1}\cos\alpha}{M_{s2}r_{s2}}\right] = 0, \tag{7c}$$

$$\ddot{U}_{c2} + \frac{C_{c2}\dot{U}_{c2}}{M_{c2}\omega_n^2} - \sum_{i=1}^{3}\frac{C_{s2n}\dot{U}_{s2pli} + C_{r2n}\dot{U}_{r2pli}}{M_{c2}\omega_n^2} + \frac{k_{c2}g\left(U_{c2}\right)}{M_{c2}\omega_n^2} - \sum_{i=1}^{3}\frac{k_{s2n}g\left(U_{s2pli}\right) + k_{r2n}g\left(U_{r2pli}\right)}{M_{c2}\omega_n^2} = 0, \tag{7d}$$

$$\ddot{U}_{s2pln} + \frac{1}{\omega_n^2}\sum_{i=1}^{3}\left[\frac{C_{s2n}\dot{U}_{s2pli}}{M_{s2}} + \frac{C_{s2n}\dot{U}_{s2pli} + C_{r2n}\dot{U}_{r2pli}}{M_{c2}}\right] + \frac{1}{\omega_n^2}\left[\frac{C_{s2n}\dot{U}_{s2pln} - C_{r2n}\dot{U}_{r2pln}}{M_{pl}} - \frac{C_{c2}\dot{U}_{c2}}{M_{c2}} - \frac{C_{c1s2}}{M_{s2}}\frac{r_{c1}\cos\alpha\dot{U}_{c1s2}}{r_{s2}}\right] + \frac{1}{\omega_n^2}\sum_{i=1}^{3}\left[\frac{k_{s2n}g\left(U_{s2pli}\right)}{M_{s2}} + \frac{k_{s2n}g\left(U_{s2pli}\right) + k_{r2n}g\left(U_{r2pli}\right)}{M_{c2}}\right] + \frac{1}{\omega_n^2}\left[\frac{k_{s2n}g\left(U_{s2pln}\right) - k_{r2n}g\left(U_{r2pln}\right)}{M_{pl}} - \frac{k_{c2}g\left(U_{c2}\right)}{M_{c2}} - \frac{k_{c1s2}}{M_{s2}} \times \frac{g\left(U_{c1s2}\right)r_{c1}\cos\alpha}{r_{s2}}\right] = 0 \quad (n = 1,2,3), \tag{7e}$$

$$\ddot{U}_{r2pln} + \frac{1}{\omega_n^2}\left[\frac{C_{r2n}\dot{U}_{r2pln} - C_{s2n}\dot{U}_{s2pln}}{M_{pl}} - \frac{C_{c2}\dot{U}_{c2}}{M_{c2}} + \frac{C_{r1r2}r_{r1}\dot{U}_{r1r2}}{M_{r2}r_{r2}}\right] + \frac{1}{\omega_n^2}\sum_{i=1}^{3}\left[\frac{C_{s2n}\dot{U}_{s2pli} + C_{r2n}\dot{U}_{r2pli}}{M_{c2}} + \frac{C_{r2n}\dot{U}_{r2pln}}{M_{r2}}\right] + \frac{1}{\omega_n^2}\left[\frac{k_{r2n}g\left(U_{r2pln}\right) - k_{s2n}g\left(U_{s2pln}\right)}{M_{pl}} - \frac{k_{c2}g\left(U_{c2}\right)}{M_{c2}} + \frac{k_{r1r2}r_{r1}g\left(U_{r1r2}\right)}{M_{r2}r_{r2}}\right]$$

$$+\frac{1}{\omega_n^2}\sum_{i=1}^{3}\left[\frac{k_{s2n}g\left(U_{s2pli}\right)+k_{r2n}g\left(U_{r2pli}\right)}{M_{c2}}+\frac{k_{r2n}g\left(U_{r2pli}\right)}{M_{r2}}\right]=\frac{-T_{\text{out}}}{M_{r2}r_{r2}b\omega_n^2}\quad(n=1,2,3),\tag{7f}$$

$$\ddot{U}_{r1r2}+\frac{C_{r1r2}\dot{U}_{r1r2}}{M_{r1r2}\omega_n^2}+\frac{1}{\omega_n^2}\sum_{i=1}^{3}\left(\frac{C_{r2n}r_{r1}}{M_{r2}r_{r2}}\dot{U}_{r2pli}-\frac{C_{r1n}}{M_{r1}}\dot{U}_{r1phi}\right)+\frac{k_{r1r2}g\left(U_{r1r2}\right)}{M_{r1r2}\omega_n^2}+\frac{1}{\omega_n^2}\sum_{i=1}^{3}\left(\frac{k_{r2n}r_{r1}g\left(U_{r2pli}\right)}{M_{r2}r_{r2}}-\frac{k_{r1n}g\left(U_{r1phi}\right)}{M_{r1}}\right)=-\frac{T_{\text{out}}r_{r1}}{M_{r2}r_{r2}^2b\omega_n^2},\tag{7g}$$

where,

$$\frac{1}{M_{s1c1}}=\frac{1}{M_{s1}}+\frac{1}{M_{c1}},\quad \frac{1}{M_{c1r1}}=\frac{1}{M_{c1}}+\frac{1}{M_{r1}},\quad \frac{1}{M_{r1r2}}=\frac{1}{M_{r1}}+\frac{r_{r1}}{M_{r2}r_{r2}}.\tag{8}$$

Equations (7a)–(7g) can be written in matrix form as

$$\ddot{\mathbf{U}}+\mathbf{C}\dot{\mathbf{U}}+\mathbf{K}g\left(\mathbf{U}\right)=\mathbf{T},\tag{9}$$

where, the piecewise-linear backlash function is defined as

$$g\left[U_{ij}\left(t\right)\right]=\begin{cases}U_{ij}\left(t\right)-b_{ij}, & U_{ij}>b_{ij},\\ 0, & -b_{ij}\le U_{ij}\le b_{ij},\\ U_{ij}\left(t\right)+b_{ij}, & U_{ij}<-b_{ij}.\end{cases}\tag{10}$$

3. Solution of Dynamic Equations

3.1. Dynamic Response Using the Numerical Integration. In this section, the mathematical model will be solved numerically by the variable step-size Runge-Kutta integration method firstly. The parameters of the closed-form planetary gear set shown in Figure 1 are given in Tables 1 and 2. In this work, the values of the gear mesh damping coefficients C_{ij} are assumed to be constants as 0.01 [11].

By resolving the dynamic differential equations of motions, the dynamic responses (displacement and speed) of low-stage carrier are gained as shown in Figure 3.

3.2. The Acquisition of Analytical Solutions. In order to establish the optimization mathematical model of the dynamic behavior of the closed-form planetary gear set, it is necessary to obtain the analytical solutions of model using the harmonic balance method. For limiting the number of algebraic balance equations, only the fundamental frequency harmonic of the mesh stiffness functions are considered in this case. Similarly, external torque functions is also considered to be in the form of mesh stiffness functions. And then, attention is paid to the periodic vibrations of system under the harmonic excitation. The procedure of solving (7a)–(7g) using HBM [5] includes some aspects as follows.

(1) According to the assumption previously mentioned, the external excitations that represent fundamental frequency pulsations can be written in the form

$$\mathbf{T}=\mathbf{T}_{mi}+\mathbf{T}_{ai}\cos\left(\Omega t+\phi_i\right),\tag{11}$$

where $\mathbf{T}_{mi}$ is the mean component of torque and $\mathbf{T}_{ai}$ is the amplitude of the alternating component of the fundamental frequency mesh force. ϕ_i is the phase angle.

(2) According to the harmonic excitations given in (11), the harmonic balance method solution to (7a)–(7g) is assumed in the same form

$$\mathbf{U}=\mathbf{U}_{mi}+\mathbf{U}_{ai}\cos\left(\Omega t+\phi_i\right),\tag{12}$$

where $\mathbf{U}_{mi}$ and $\mathbf{U}_{ai}$ are the mean and alternating components of the steady state response, respectively, and ϕ_i is the phase angle.

(3) For relative mesh displacements, the piecewise-linear function in (4a)–(4h) can be written in a unified form:

$$g\left(U_i\right)=N_{mi}U_{mi}+N_{ai}U_{ai}\cos\left(\Omega t+\phi_i\right),\tag{13}$$

where

$$N_{mi}=1+\frac{q_{ai}}{2q_{mi}}\left[G\left(\mu_+\right)-G\left(\mu_-\right)\right],\quad N_{ai}=1-\frac{1}{2}\left[H\left(\mu_+\right)-H\left(\mu_-\right)\right].\tag{14}$$

The expressions of G and H are given in the Appendix.

(4) Considering the mean value and the fundamental harmonic value of periodically time-varying mesh stiffness in Figure 2, the elements of stiffness matrix $\mathbf{K}$ in (9) can be written as

$$k_{ij}=k_{mij}+k_{aij}\cos\left(\Omega\tau+\varphi_{ij}\right).\tag{15}$$

So, the stiffness matrix $\mathbf{K}$ is written in terms of two separate matrices for mean stiffness and alternating stiffness as

$$\mathbf{K}=\mathbf{K}_m+\Delta\mathbf{K},\tag{16}$$

where

$$\mathbf{K}_m=\left[k_{mij}\right]_{n\times n},\quad \Delta\mathbf{K}=\left[k_{aij}\cos\left(\Omega\tau+\varphi_{ij}\right)\right]_{n\times n}.\tag{17}$$

Table 1: Geometric parameters of the closed-form planetary gear set.

	High-speed stage			Low-speed stage		
	Sun gear	Planet gear	Ring gear	Sun gear	Planet gear	Ring gear
Number of teeth	16	38	92	16	24	62
Module (mm)		3			3	
Pressure angle (deg)		21.5°			21.5°	
Modification coefficient	0.3	0.12	0.54	0.475	0.471	0.309
Face width (mm)	22	20	20	34.5	30	32

Table 2: Physical parameters of the closed-form planetary gear set.

	High-speed stage				Low-speed stage			
	Sun	Planet	Ring	Carrier	Sun	Planet	Ring	Carrier
Inertia (kg·mm^2)	88.54	827	68883.72	6800.79	339.59	462.92	110213.96	9400.9
Rotation radius (mm)	11.07	32.78	113	57.25	23.28	31.12	113	28
Mean mesh stiffness (N/mm)	$K_{s1n} = K_{r1n} = 2e^7$				$K_{s2n} = K_{r2n} = 2e^7$			
Torsional stiffness (N/mm)	$K_{c1s2} = 2e^7$				$K_{r1r2} = 10e^7$, $K_{c2} = 10e^8$			
Total mass	69 kg							

By substituting (9)–(14) into (7a)–(7g) and balancing the like harmonic terms, the algebraic equations of system can be gained

$$\mathbf{K}_m\mathbf{y}_m + \frac{\mathbf{K}_1\mathbf{y}_3}{2} + \frac{\mathbf{K}_2\mathbf{y}_4}{2} - \mathbf{T}_m = \{0\}_{n\times1},$$
$$\mathbf{K}_m\mathbf{y}_3 + \mathbf{K}_1\mathbf{y}_m - \Omega^2\mathbf{y}_1 - \Omega\mathbf{C}\mathbf{y}_2 - \mathbf{T}_1 = \{0\}_{n\times1}, \quad (18)$$
$$\mathbf{K}_m\mathbf{y}_4 + \mathbf{K}_2\mathbf{y}_m - \Omega^2\mathbf{y}_2 + \mathbf{C}\Omega\mathbf{y}_1 - \mathbf{T}_2 = \{0\}_{n\times1},$$

where $\mathbf{y}_1 = \{q_{ai}\cos\varphi_i\}_{n\times1}$, $\mathbf{y}_2 = \{q_{ai}\sin\varphi_i\}_{n\times1}$, $\mathbf{y}_3 = \{N_{ai}q_{ai}\cos\varphi_i\}_{n\times1}$, $\mathbf{y}_4 = \{N_{ai}q_{ai}\sin\varphi_i\}_{n\times1}$, $\mathbf{y}_m = \{N_{mi}q_{mi}\}_{n\times1}$, $\mathbf{T}_1 = \{T_{ai}\cos\phi_i\}_{n\times1}$, $\mathbf{T}_2 = \{T_{ai}\sin\phi_i\}_{n\times1}$, $\mathbf{K}_1 = \{k_{aij}\cos\phi_{ij}\}_{n\times1}$, and $\mathbf{K}_2 = \{k_{aij}\sin\phi_{ij}\}_{n\times1}$.

The matrices $\mathbf{k}_m$ and $\mathbf{C}$ are given in the Appendix.

4. Optimization of the Transmission System

4.1. Variables and Objective Function of Optimization. A number of key design parameters have great influence on the dynamic characteristics of the gear transmissions, so they can be chosen as design variables, such as number of each gear, module, gear width, and pressure angle. To simplify optimization process, the module of each stage m_1, m_2, number of sun gears of each stage z_{s1}, z_{s2}, and pressure angle of each stage α_1, α_2 are chosen as optimization variables in this study. So, the recurrence variables vector for this optimization procedure can be written as

$$x = [m_1, m_2, z_{s1}, z_{s2}, \alpha_1, \alpha_2]^T. \quad (19)$$

The dynamic characteristics standards of the gear transmission include maximum dynamic loads, dynamic load factor, stiffness, and displacement/velocity/acceleration of vibration, each of these standards can be chosen as optimization objective. Considering that the rotation center of the planets in low-stage is unfixed, and reducing the total weight of the transmission system simultaneously, the displacement of rotational vibration of the low-stage carrier and total mass of the gear set are chosen as optimization objective. According to the basic idea of the multiobjective optimization, two optimization objective functions unified using the normalized weighting method [11] can be written in form

$$f = \lambda_1 f_1 + \lambda_2 f_2, \quad (20)$$

where λ_1 and λ_2 are weighting coefficient, here, $\lambda_1 = 0.3$ and $\lambda_2 = 0.7$ [12]. f_1 and f_2 are the ampler of the rotational vibration displacement of the low-stage carrier and total mass of the gear set, respectively. Due to limited space, the expressions of f_1 and f_2 are not given in detail.

4.2. Constraints

(1) Distributing the number of teeth of each gear: concentric conditions, adjacent conditions, and assembly conditions;

(2) contact fatigue strength and bending fatigue strength constraints

$$S_H(i) \ge S_{H\min}, \qquad S_F(i) \ge S_{F\min}, \qquad i = s_1, s_2, r_1, r_2, ph, pl, \quad (21)$$

(3) contact ratio constraints

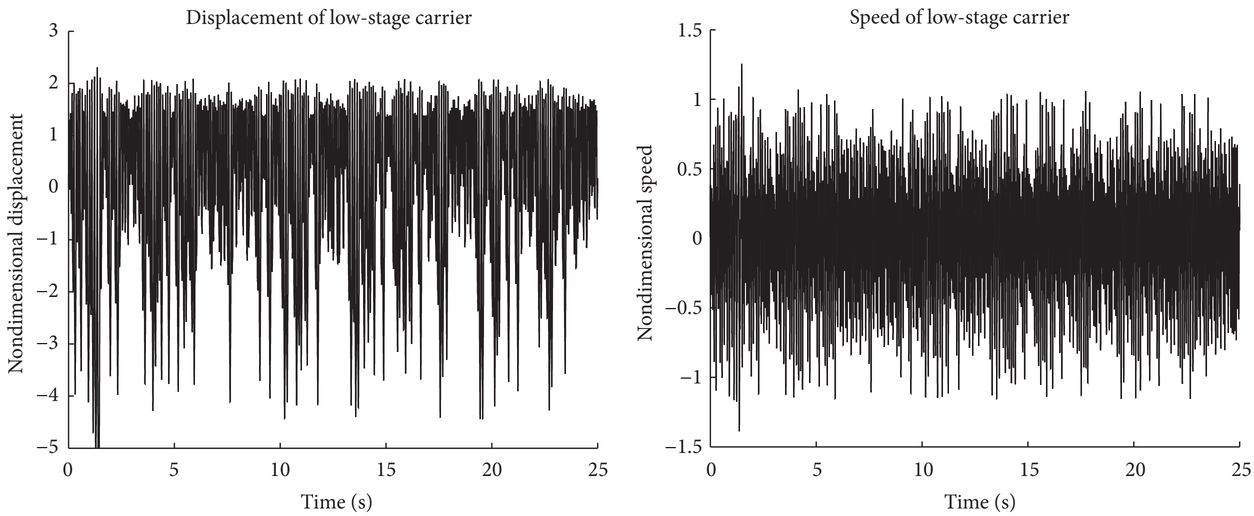
$$\varepsilon_{ij} \ge 1.5, \quad i = s_1, s_2, r_1, r_2, \quad j = ph, pl, \quad (22)$$

(4) transmission ratio without loop-power conditions

$$-32 \le i_1 + i_2 - i_1 i_2 \le -30,$$
$$0 < \frac{i_2 - i_1 i_2}{i_1 + i_2 - i_1 i_2} < 1, \quad (23)$$

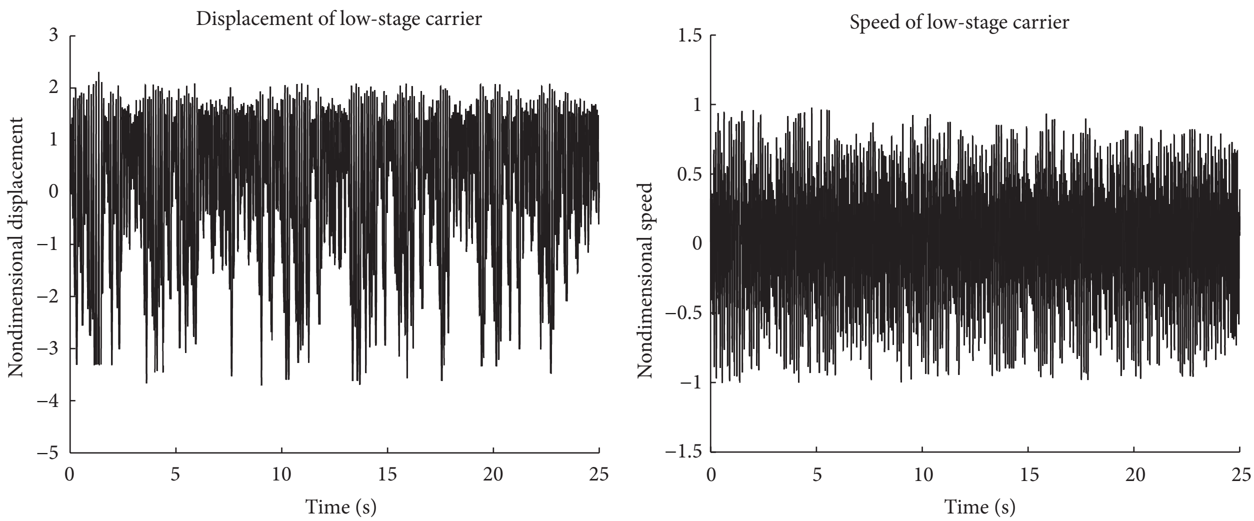
Table 3: Parameters of structure after optimization.

	High-speed stage			Low-speed stage		
	Sun gear	Planet gear	Ring gear	Sun gear	Planet gear	Ring gear
Number of teeth	17	40	97	18	23	66
Module (mm)		2			3	
Pressure angle (deg)		20°			20°	
Total mass (kg)			63.5 kg			



(a) Displacement of low-speed stage carrier (b) Speed of low-speed stage carrier

Figure 3: Response of low-speed stage carrier without optimization.



(a) Displacement of low-speed stage carrier (b) Speed of low-speed stage carrier

Figure 4: Response of low-speed stage carrier after optimization.

where

$$i_n = -\frac{Z_{rn}}{Z_{sn}}, \quad n = 1, 2, \tag{24}$$

(5) minimum tooth thickness constraints

$$S_{ai} \geq 0.4 m_j, \quad i = s_1, s_2, r_1, r_2, ph, pl; \;\; j = 1, 2. \tag{25}$$

5. Results and Discussions

Some more logical parameters of the closed-form planetary gear set are obtained according to optimization constraints conditions previously mentioned: $m_1 = 2$, $m_2 = 3$, $Z_{s1} = 17$, $Z_{s2} = 18$, $\alpha_1 = \alpha_2 = 20^\circ$. Further, other parameters of transmission can be computed as shown in Table 3. Total mass of gear set is 63.5 kg. Comparing the results between before and after the optimization, it is clear that transmission system reduces the total mass of 8.6% and the dynamic characteristic significantly improves as well. It is also evident from the comparison of Figure 3 with Figure 4 that the displacement of each structure significantly reduces at the same time. It is obvious that the scope of the amplitude of displacement of low-stage carrier is −3 to 3 in Figure 4(a), while that in Figure 3(a) is −5 to 3.

6. Conclusions

Considering the gear backlash, time-varying mesh stiffness and excitation fluctuation, and so forth, a discrete nonlinear dynamic model of a two stages closed-form planetary set was proposed in this study. In order to facilitate the analysis and comparison of system response between before and after optimization, nonlinear differential equations of motion of the dynamic model were solved using a Runge-Kutta numerical integration method. For optimization of transmission system, the analytical solutions were obtained. The total optimization objective function is obtained using the normalized weights method.

A case shows that the total mass of transmission system has significantly reduced and dynamic characteristic has distinctly improved. Effectiveness of the dynamic model and the dynamic optimization mathematic model is demonstrated.

Appendix

Consider the following:

$$G(\mu) = \begin{cases} \left(\dfrac{2}{\pi}\right)\left(\mu \sin^{-1}\mu + \sqrt{1-\mu^2}\right), & |\mu| \le 1, \\ |\mu|, & |\mu| > 1, \end{cases}$$
$$H(\mu) = \begin{cases} -1, & \mu < -1, \\ \left(\dfrac{2\mu}{\pi}\right)\left(\sin^{-1}\mu + \sqrt{1-\mu^2}\right), & |\mu| \le 1, \\ 1, & \mu > 1. \end{cases} \tag{A.1}$$

The elements of the stiffness matrix $\mathbf{k}_m$ are given as follows

$$k(1,1) = \frac{k_{s1n}}{\omega_n^2}\left(\frac{1}{M_{s1}} + \frac{1}{M_{c1}} + \frac{1}{M_{ph}}\right),$$
$$k(1,i) = \frac{k_{s1n}}{\omega_n^2}\left(\frac{1}{M_{s1}} + \frac{1}{M_{c1}}\right), \quad i = 2,3,$$
$$k(1,4) = \frac{k_{r1n}}{\omega_n^2}\left(\frac{1}{M_{c1}} - \frac{1}{M_{ph}}\right),$$
$$k(1,i) = \frac{k_{r1n}}{M_{c1}\omega_n^2}, \quad i = 5,6,$$
$$k(1,7) = -\frac{k_{c1s2}}{M_{c1}\omega_n^2},$$
$$k(2,i) = \frac{k_{s1n}}{\omega_n^2}\left(\frac{1}{M_{s1}} + \frac{1}{M_{c1}}\right), \quad i = 1,3,$$
$$k(2,7) = -\frac{k_{c1s2}}{M_{c1}\omega_n^2},$$
$$k(2,2) = \frac{k_{s1n}}{\omega_n^2}\left(\frac{1}{M_{s1}} + \frac{1}{M_{c1}} + \frac{1}{M_{ph}}\right),$$
$$k(2,i) = \frac{k_{r1n}}{M_{c1}\omega_n^2}, \quad i = 4,6,$$
$$k(2,5) = \frac{k_{r1n}}{\omega_n^2}\left(\frac{1}{M_{c1}} - \frac{1}{M_{ph}}\right),$$
$$k(2,i) = \frac{k_{s1n}}{\omega_n^2}\left(\frac{1}{M_{s1}} + \frac{1}{M_{c1}}\right), \quad i = 1,3,$$
$$k(2,2) = \frac{k_{s1n}}{\omega_n^2}\left(\frac{1}{M_{s1}} + \frac{1}{M_{c1}} + \frac{1}{M_{ph}}\right),$$
$$k(2,i) = \frac{k_{r1n}}{M_{c1}\omega_n^2}, \quad i = 4,6,$$
$$k(2,5) = \frac{k_{r1n}}{\omega_n^2}\left(\frac{1}{M_{c1}} - \frac{1}{M_{ph}}\right),$$
$$k(2,7) = -\frac{k_{c1s2}}{M_{c1}\omega_n^2},$$
$$k(3,i) = \frac{k_{s1n}}{\omega_n^2}\left(\frac{1}{M_{s1}} + \frac{1}{M_{c1}}\right), \quad i = 1,2,$$
$$k(3,3) = \frac{k_{s1n}}{\omega_n^2}\left(\frac{1}{M_{s1}} + \frac{1}{M_{c1}} + \frac{1}{M_{ph}}\right),$$
$$k(3,i) = \frac{k_{r1n}}{M_{c1}\omega_n^2}, \quad i = 4,5,$$
$$k(3,6) = \frac{k_{r1n}}{\omega_n^2}\left(\frac{1}{M_{c1}} - \frac{1}{M_{ph}}\right),$$
$$k(3,7) = -\frac{k_{c1s2}}{M_{c1}\omega_n^2},$$
$$k(4,1) = \frac{k_{s1n}}{\omega_n^2}\left(\frac{1}{M_{c1}} - \frac{1}{M_{ph}}\right),$$
$$k(4,i) = \frac{k_{s1n}}{M_{c1}\omega_n^2}, \quad i = 2,3,$$
$$k(4,7) = -\frac{k_{c1s2}}{M_{c1}\omega_n^2},$$
$$k(4,4) = \frac{k_{r1n}}{\omega_n^2}\left(\frac{1}{M_{ph}} + \frac{1}{M_{c1}} + \frac{1}{M_{r1}}\right),$$

$$k(4,i) = \frac{k_{r1n}}{\omega_n^2}\left(\frac{1}{M_{c1}} + \frac{1}{M_{r1}}\right), \quad i = 5, 6,$$

$$k(4,15) = -\frac{k_{r1r2}}{M_{r1}\omega_n^2},$$

$$k(5,i) = \frac{k_{s1n}}{M_{c1}\omega_n^2}, \quad i = 1, 3,$$

$$k(5,2) = \frac{k_{s1n}}{\omega_n^2}\left(\frac{1}{M_{c1}} - \frac{1}{M_{ph}}\right),$$

$$k(5,i) = \frac{k_{r1n}}{\omega_n^2}\left(\frac{1}{M_{c1}} + \frac{1}{M_{r1}}\right), \quad i = 5, 6,$$

$$k(5,5) = \frac{k_{r1n}}{\omega_n^2}\left(\frac{1}{M_{ph}} + \frac{1}{M_{c1}} + \frac{1}{M_{r1}}\right),$$

$$k(5,7) = -\frac{k_{c1s2}}{M_{c1}\omega_n^2},$$

$$k(5,15) = -\frac{k_{r1r2}}{M_{r1}\omega_n^2},$$

$$k(6,i) = \frac{k_{s1n}}{M_{c1}\omega_n^2}, \quad i = 1, 2,$$

$$k(6,3) = \frac{k_{s1n}}{\omega_n^2}\left(\frac{1}{M_{c1}} - \frac{1}{M_{ph}}\right),$$

$$k(6,i) = \frac{k_{r1n}}{\omega_n^2}\left(\frac{1}{M_{c1}} + \frac{1}{M_{r1}}\right), \quad i = 4, 5,$$

$$k(6,6) = \frac{k_{r1n}}{\omega_n^2}\left(\frac{1}{M_{ph}} + \frac{1}{M_{c1}} + \frac{1}{M_{r1}}\right),$$

$$k(6,7) = -\frac{k_{c1s2}}{M_{c1}\omega_n^2},$$

$$k(6,15) = -\frac{k_{r1r2}}{M_{r1}\omega_n^2},$$

$$k(7,i) = -\frac{k_{s1n}}{M_{c1}\omega_n^2}, \quad i = 1, 2, 3,$$

$$k(7,i) = -\frac{k_{r1n}}{M_{c1}\omega_n^2}, \quad i = 4, 5, 6,$$

$$k(7,7) = \frac{k_{c1s2}}{\omega_n^2}\left[\frac{1}{M_{c1}} + \frac{1}{M_{s2}}\left(\frac{r_{c1}\cos\alpha}{r_{s2}}\right)^2\right],$$

$$k(7,i) = -\frac{k_{s2n}}{M_{s2}\omega_n^2}\frac{r_{c1}\cos\alpha}{r_{s2}}, \quad i = 9, 10, 11,$$

$$k(8,8) = \frac{k_{c2}}{M_{c2}\omega_n^2},$$

$$k(8,i) = -\frac{k_{s2n}}{M_{c2}\omega_n^2}, \quad i = 9, 10, 11,$$

$$k(8,i) = -\frac{k_{r2n}}{M_{c2}\omega_n^2}, \quad i = 12, 13, 14,$$

$$k(9,7) = -\frac{k_{c1s2}}{M_{s2}\omega_n^2}\frac{r_{c1}\cos\alpha}{r_{s2}},$$

$$k(9,8) = -\frac{k_{c2}}{M_{c2}\omega_n^2},$$

$$k(9,9) = \frac{k_{s2n}}{\omega_n^2}\left(\frac{1}{M_{s2}} + \frac{1}{M_{c2}} + \frac{1}{M_{pl}}\right),$$

$$k(9,i) = \frac{k_{s2n}}{\omega_n^2}\left(\frac{1}{M_{s2}} + \frac{1}{M_{c2}}\right), \quad i = 10, 11,$$

$$k(9,12) = \frac{k_{r2n}}{\omega_n^2}\left(\frac{1}{M_{c2}} - \frac{1}{M_{pl}}\right),$$

$$k(9,i) = \frac{k_{r2n}}{M_{c2}\omega_n^2}, \quad i = 13, 14,$$

$$k(10,7) = -\frac{k_{c1s2}}{M_{s2}\omega_n^2}\frac{r_{c1}\cos\alpha}{r_{s2}},$$

$$k(10,8) = -\frac{k_{c2}}{M_{c2}\omega_n^2},$$

$$k(10,i) = \frac{k_{s2n}}{\omega_n^2}\left(\frac{1}{M_{s2}} + \frac{1}{M_{c2}}\right), \quad i = 9, 10,$$

$$k(10,10) = \frac{k_{s2n}}{\omega_n^2}\left(\frac{1}{M_{s2}} + \frac{1}{M_{c2}} + \frac{1}{M_{pl}}\right),$$

$$k(10,i) = \frac{k_{r2n}}{M_{c2}\omega_n^2}, \quad i = 12, 14,$$

$$k(10,13) = \frac{k_{r2n}}{\omega_n^2}\left(\frac{1}{M_{c2}} - \frac{1}{M_{pl}}\right),$$

$$k(11,7) = -\frac{k_{c1s2}}{M_{s2}\omega_n^2}\frac{r_{c1}\cos\alpha}{r_{s2}},$$

$$k(11,8) = -\frac{k_{c2}}{M_{c2}\omega_n^2},$$

$$k(11,i) = \frac{k_{s2n}}{\omega_n^2}\left(\frac{1}{M_{s2}} + \frac{1}{M_{c2}}\right), \quad i = 9, 10,$$

$$k(11,11) = \frac{k_{s2n}}{\omega_n^2}\left(\frac{1}{M_{s2}} + \frac{1}{M_{c2}} + \frac{1}{M_{pl}}\right),$$

$$k(11,i) = \frac{k_{r2n}}{M_{c2}\omega_n^2}, \quad i = 12, 13,$$

$$k(11,14) = \frac{k_{r2n}}{\omega_n^2}\left(\frac{1}{M_{c2}} - \frac{1}{M_{pl}}\right),$$

$$k(12,8) = -\frac{k_{c2}}{M_{c2}\omega_n^2},$$

$$k(12,15) = \frac{k_{r1r2} r_{r1}}{M_{r2}\omega_n^2 r_{r2}},$$

$$k(12,9) = \frac{k_{s2n}}{\omega_n^2}\left(\frac{1}{M_{c2}} - \frac{1}{M_{pl}}\right),$$

$$k(12,i) = \frac{k_{s2n}}{M_{c2}\omega_n^2}, \quad i = 10, 11,$$

$$k(12,12) = \frac{k_{r2n}}{\omega_n^2}\left(\frac{1}{M_{pl}} + \frac{1}{M_{c2}}\right),$$

$$k(12,i) = \frac{k_{r2n}}{M_{c2}\omega_n^2}, \quad i = 13, 14,$$

$$k(13,8) = -\frac{k_{c2}}{M_{c2}\omega_n^2},$$

$$k(13,10) = \frac{k_{s2n}}{\omega_n^2}\left(\frac{1}{M_{c2}} - \frac{1}{M_{pl}}\right),$$

$$k(13,i) = \frac{k_{s2n}}{M_{c2}\omega_n^2}, \quad i = 9, 11,$$

$$k(13,13) = \frac{k_{r2n}}{\omega_n^2}\left(\frac{1}{M_{pl}} + \frac{1}{M_{c2}}\right),$$

$$k(13,i) = \frac{k_{r2n}}{M_{c2}\omega_n^2}, \quad i = 12, 14,$$

$$k(13,15) = \frac{k_{r1r2} r_{r1}}{M_{r2}\omega_n^2 r_{r2}},$$

$$k(14,8) = -\frac{k_{c2}}{M_{c2}\omega_n^2},$$

$$k(14,11) = \frac{k_{s2n}}{\omega_n^2}\left(\frac{1}{M_{c2}} - \frac{1}{M_{pl}}\right),$$

$$k(14,i) = \frac{k_{s2n}}{M_{c2}\omega_n^2}, \quad i = 9, 10,$$

$$k(14,14) = \frac{k_{r2n}}{\omega_n^2}\left(\frac{1}{M_{pl}} + \frac{1}{M_{c2}}\right),$$

$$k(14,i) = \frac{k_{r2n}}{M_{c2}\omega_n^2}, \quad i = 12, 13,$$

$$k(14,15) = \frac{k_{r1r2} r_{r1}}{M_{r2}\omega_n^2 r_{r2}},$$

$$k(15,i) = -\frac{k_{r1n}}{M_{r1}\omega_n^2}, \quad i = 4, 5, 6,$$

$$k(15,i) = \frac{k_{r2n} r_{r1}}{M_{r2} r_{r2}\omega_n^2}, \quad i = 12, 13, 14,$$

$$k(15,15) = \frac{k_{r1r2}}{\omega_n^2}\left(\frac{1}{M_{r1}} + \frac{r_{r1}}{M_{r2} r_{r2}}\right). \tag{A.2}$$

The elements of $\mathbf{k}_m$ not listed previously are taken as zeros. The matrix $\mathbf{C}$ is in the similar form of $\mathbf{k}_m$ and no longer is listed in detail here.

Notations

HBM: Harmonic balance method
b: Half of clearance (backlash)
k: Gear mesh stiffness
r: Ring gear
g: Discontinuous displacement function
I: Polar mass moment of inertia
i: Transmission ratio
c: Carrier
m: Actual mass
$\bar{t}$: Dimensional time
u: Actual equivalent transverse displacement in the direction of pressure line
U: Relative equivalent transverse displacement in the direction of pressure line
s: Sun gear
n_p: Number of planet
$\mathbf{T}$: Torque
θ : Angular displacement
C: Damping coefficient
Z: Number of gear
r: Base cycle radius
M: Equivalent mass
t: Nondimensional time
α: Pressure angle.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (no. 51175299), Shandong Province Natural Science Foundation of China (ZR2010EM012) and Shandong University Independent Innovation Foundation of China (no. 2012TS048).

References

[1] F. Cunliffe, J. D. Smith, and D. B. Welbourn, "Dynamic tooth loads in epicyclic gears," *Transactions of the ASME Journal of Engineering for Industry*, vol. 96, no. 2, pp. 578–584, 1974.
[2] M. Botman, "Epicyclic gear vibrations," *Transactions of the ASME Journal of Engineering for Industry*, vol. 98, no. 3, pp. 811–815, 1976.
[3] G. Antony, "Gear vibration-investigation of the dynamic behavior of one stage epicyclic gears," AGMA Technical Paper 88-FTM-12, 1988.
[4] A. Kahraman, "Free torsional vibration characteristics of compound planetary gear sets," *Mechanism and Machine Theory*, vol. 36, no. 8, pp. 953–971, 2001.
[5] T. Sun and H. Hu, "Nonlinear dynamics of a planetary gear system with multiple clearances," *Mechanism and Machine Theory*, vol. 38, no. 12, pp. 1371–1390, 2003.
[6] A. Al-shyyab and A. Kahraman, "A non-linear dynamic model for planetary gear sets," *Proceedings of the Institution of Mechanical Engineers K: Journal of Multi-body Dynamics*, vol. 221, no. 4, pp. 567–576, 2007.

[7] A. Al-Shyyab, K. Alwidyan, A. Jawarneh, and H. Tlilan, "Non-linear dynamic behaviour of compound planetary gear trains: model formulation and semi-analytical solution," *Proceedings of the Institution of Mechanical Engineers K: Journal of Multi-body Dynamics*, vol. 223, no. 3, pp. 199–210, 2009.

[8] T. Zeng Bao, Z. Yi Fang, and L. Wei Zhong, "Many stage helical gear train system optimal design which takes the dynamic properties as an objective function," *Chinese Journal of Mechanical Engineering*, vol. 30, no. 5, pp. 66–75, 1994 (Chinese).

[9] X. Guan Wei, Y. Jia Jun, X. LaiYuan et al., "Dynamic optimal design on planetary driver," *Mechanical Science and Technology*, vol. 17, no. 2, pp. 209–217, 1998 (Chinese).

[10] G. W. Blankenship and A. Kahraman, "Steady state forced response of a mechanical oscillator with combined parametric excitation and clearance type non-linearity," *Journal of Sound and Vibration*, vol. 185, no. 5, pp. 743–765, 1995.

[11] H. Guang and W. Gui Cheng, "Dynamics optimization design of closed epicyclic gear trains," *Mechanical Science and Technology*, vol. 23, no. 9, pp. 1131–1134, 2004 (Chinese).

[12] D.-T. Qin, X.-G. Gu, J.-H. Wang, and J.-G. Liu, "Dynamic analysis and optimization of gear trains in a megawatt level wind turbine," *Journal of Chongqing University*, vol. 32, no. 4, pp. 408–414, 2009 (Chinese).

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 910209, 15 pages
http://dx.doi.org/10.1155/2013/910209

Research Article

Characterization of the Transient Response of Coupled Optimization in Multidisciplinary Design

Erich Devendorf[1] and Kemper Lewis[2]

[1] *Air Force Research Laboratory, Information Directorate, Rome, NY 13044, USA*
[2] *Department of Mechanical and Aerospace Engineering, University at Buffalo—SUNY, Buffalo, NY 14260, USA*

Correspondence should be addressed to Kemper Lewis; kelewis@buffalo.edu

Received 11 February 2013; Accepted 7 May 2013

Academic Editor: Dan Zhang

Time is an asset of critical importance in a multidisciplinary design process and it is desirable to reduce the amount of time spent designing products and systems. Design is an iterative activity and designers consume a significant portion of the product development process negotiating a mutually acceptable solution. The amount of time necessary to complete a design depends on the number and duration of design iterations. This paper focuses on accurately characterizing the number of iterations required for designers to converge to an equilibrium solution in distributed design processes. In distributed design, systems are decomposed into smaller, coupled design problems where individual designers have control over local design decisions and seek to achieve their own individual objectives. These smaller coupled design optimization problems can be modeled using coupled games and the number of iterations required to reach equilibrium solutions varies based on initial conditions and process architecture. In this paper, we leverage concepts from game theory, classical controls, and discrete systems theory to evaluate and approximate process architectures without carrying out any solution iterations. As a result, we develop an analogy between discrete decisions and a continuous time representation that we analyze using control theoretic techniques.

1. Introduction

The design of complex systems presents both technical and logistical challenges to organizations. In some cases, organizations do not have the requisite technical expertise to overcome design challenges or meet design requirements. However, even when an organization possesses sufficient technical expertise, there are many cases where inadequate logistical control can inhibit the application of this design expertise in a meaningful manner. The logistical challenges facing organizations are becoming increasingly significant as the size and sophistication of modern engineered products increase. This growth in sophistication and size leads to the decomposition of design systems as a means to reduce system complexity. Although system decomposition reduces the technical complexity of each subsystem's individual design problems by reducing their size, it introduces significant logistical challenges. Also, it often requires significant approximation of nonlocal behavior, interfaces, and solutions. These decomposed systems are often large and multidisciplinary in nature, with diverse subsystems governed by unique objectives and constraints.

To effectively manage decomposed design problems, it is important to understand how their constituent subsystems interact with one another. One approach to capture and analyze these interactions is to model the collection of subsystems as a distributed design process. In distributed design, understanding subsystem interactions is fundamental to predicting the system equilibrium properties and transient response. To predict these systems' equilibrium properties, which include location and stability, a game theoretic approach was introduced in [1]. This analysis was further supported using an analogy to the cobweb model and proven using mathematical induction in [2]. For large systems, a discrete time linear system approach was used in [3] and was generalized to handle different process configurations in [4, 5]. In this paper, the linear system model developed in [6] is used as the basis for the approximation and analysis of the transient response of distributed design processes.

Transient response refers to the dynamic behavior of a distributed design system, beginning with the first design iteration and ending when the subsystems reach equilibrium. The transient response possesses two properties: (1) the convergence shape and (2) the convergence time [7]. The convergence shape depends on the properties of the distributed design process and could be sinusoidal, exponential, or some combination of responses. Examining the convergence shape is a topic of future work. This paper focuses on convergence time and develops an approach to estimate an upper bound for the number of design iterations required for a distributed design process to converge to an equilibrium solution.

Other work has been done to characterize and approximate the solutions to certain types of two-person games including bimatrix games characterized by nonsymmetrical fuzzy approximations [8–10], multiplayer congestion games [11, 12], and discontinuous games [13]. However, in this work, we do not focus on the approximation of the solution to particular games, but rather the convergence process to arrive at such a solution, if it exists. We focus on two-player continuous games, as they provide a characterization of coupled optimization problems frequently occurring in complex design problems.

The intellectual merit of this work is based on its insight into the dynamics of distributed design processes. The linear system model used in previous work to assess stability is refined to analyze both system stability and transient response. Through the refinement of this model, it is demonstrated that the uniquely discontinuous nature of any decentralized decision network must be considered when they are modeled using systems theory. Furthermore, several control theory principles are shown to be valid when analyzing distributed design processes.

As an initial context for this investigation, we focus on distributed design problems with unconstrained quadratic objective functions. These types of problems are well modeled using linear system theory, but it is recognized that many design problems cannot easily be converted into unconstrained problems and/or do not have quadratic objective functions. While techniques in metamodeling exist to represent higher-order systems using quadratic response surfaces [14], it is desirable to analyze systems in their native mathematical form. We examine quadratic systems in this paper in order to fully understand their fundamental principles before applying the concepts to higher-order systems. As a result, this work represents a critical first step towards leveraging linear system analysis techniques to understand and analyze the behavior of distributed design systems.

This work has broader impacts in any scenario where the decentralization of decisions is present. These scenarios can range from product design to coordinating disaster relief. It provides a deeper understanding of the decision-making process and enables a greater level of process control. This enables decision makers to reach iterative solutions quicker and to set realistic deadlines or timetables. It represents a unique and effective approach to find an upper bound for the time it takes a distributed design process to converge to a stable equilibrium. Furthermore, it provides insight into how to best configure these processes to minimize the maximum number of iterations required to reach equilibrium.

The following sections provide background into concepts foundational to the examination of decentralized decision networks. In Section 2 an overview of stability and convergence concerns in multidisciplinary optimization is presented along with the basic tenets of distributed design processes. Based on the tenets outlined in Section 2, a linear systems approach to examine the transient response of distributed design systems is examined in Section 3. Finally, these results are summarized and areas of future work are identified in Section 4.

2. Materials and Methods

MDO problems have two classifications based on the process used to complete the design [15]. From a structural perspective, the simplest method to solve an MDO problem is to apply an all-at-once approach. In an all-at-once approach designers from different disciplines work as system analyzers to determine objective function and constraint values for a single optimization problem [16, 17]. There are significant advantages to organizing an MDO system to solve a single centralized optimization problem. In a centralized problem, all designers are working towards the exact same objective and information about the entire system is available to designers.

Although these advantages make centralization attractive, it is almost impossible to centralize the design of complex systems. Three major design approaches, Systematic Design, Total Design, and Axiomatic Design, are all structured with the understanding that some level of system decomposition is often desirable to speed development times through parallelization, to reduce system complexity, and to reduce computational time [18–20]. Recognizing this, it is likely that decomposition will remain an important and necessary aspect of product design processes in the near future.

Fortunately for designers a wide range of approaches have been developed to aid in system decomposition. The system decomposition process can be broken into two fundamental steps: (1) identify the necessary subsystems, (2) establish a framework to govern subsystem interactions. The first step in this process is not the topic of this paper, but it is by no means a trivial task. Subsystems can be created based on object decomposition, aspect decomposition, sequential decomposition, and model-based decomposition [21]. A survey of the relative merits of these subsystem creation approaches was performed by Sobieszczanski-Sobieski and Haftka [15].

The second step in system decomposition is the creation of a design framework. MDO frameworks specify the mechanics of how the design problem is solved including the subsystem objective functions, communication protocols, design variable control, and the other coordination procedures required for the subsystems to effectively iterate to a solution. There are a wide range of MDO frameworks which handle these issues differently while making certain guarantees about system convergence and the optimality of the final converged solution. These approaches include analytic target cascading [22], concurrent subspace optimization [23, 24], bilevel integrated system synthesis [25] and collaborative optimization [26].

Each of these frameworks has its own advantages. For example, analytic target cascading guarantees that the decentralized system converges and that the converged value is globally optimal [27]. It also provides for traceability and facilitates the integration of marketing, business, and design systems [28, 29].

There are several reasons why an MDO framework may not be applied to a complex design problem. Applying a framework requires a significant level of coordination between subsystems and a high level of management expertise [15]. Furthermore, the engineering and design personnel involved must all agree to some extent to the proposed decomposition and framework. There are also some cases that do not naturally lend themselves to formal decomposition or where the parties involved cannot agree on an appropriate framework. In these cases subsystems often act exclusively in their own self interest, attempting to most effectively solve their individual optimization formulations, and communication between subsystems is dictated by the required exchange of design information. When these conditions exist, design problems can be well modeled as distributed design processes. Even when there is communication between design teams, a distributed design process can often be used to model these systems. The assumptions and mechanics governing distributed design problems are discussed in Section 2.1.

2.1. Conditions for Distributed Design Processes. Distributed design processes are iterative and can be cooperative, noncooperative, or a hybrid of the two. This work examines noncooperative distributed design processes where design subsystems often have conflicting objectives or organizational barriers that prevent them from fully cooperating synchronously. Even when subsystems share the same aggregate design goals, there are cases where they will compete with one another for design resources [30] due to the underlying scarcity of such resources.

In addition to being categorized based on cooperation, distributed design systems can be broken into hierarchical and non-hierarchical realizations. In this work, it is assumed that a design process can be adequately modeled as a non-hierarchical process. This assumption is not overly restrictive in scenarios where distributed design processes are applied, since these scenarios typically lack a strong system level presence. We elaborate the non-cooperative protocol used in this paper in Section 2.2 and discuss criteria for equilibrium stability and transient response in Sections 2.3 and 2.4, respectively. The concept of solution process architecture is introduced in Section 2.5.

2.2. Criteria for Noncooperation in Repeated Games. Efforts to model and analyze subsystem interactions have led to the development of many different models for distributed design processes. One of the first models for distributed design applied mathematical notions of game theory to model design subsystems as players in a non-cooperative game. These players act independently of one another and through successive plays of the game, they eventually reach an equilibrium solution [31]. This model forms the basis for the analysis of distributed design systems in this work and is foundational to the application of other distributed design process models. These assumptions have been defined for distributed design in [2] and are as follows.

(1) Subsystems have knowledge of only their own local objectives;

(2) Subsystems act unilaterally in accordance with their own objectives;

(3) Subsystems have complete control over specific local design variables;

(4) Subsystems communicate by sharing the current value of their local design variables.

The applicability of these assumptions to decentralized design problems is discussed in various contexts in [32–35] using examples that include the design of passenger aircraft, automotive engines, semiconductor chips, and steam turbines. Distributed design problems can also emerge as iterative subproblems in a larger MDO process. For example, in [36, 37] the ordering of decomposed design systems was examined and iterative loops emerged due to subsystem coupling of concurrently executed tasks.

Subsystems in distributed design processes have their own specific objectives they are attempting to achieve. They have complete control over a set of local design variables which appear in their own objective formulation. These variables also act to couple subsystems in a manner that restricts the ability of a subsystem to independently achieve its objectives. The influence of this coupling has been investigated using network theory to model distributed design processes [38]. The examination of such coupling using game theory to define, identify, and classify system equilibrium is the focus of Section 2.3.

2.3. Equilibrium Stability for Noncooperative Processes. Determining if a design system converges to a stable solution is of critical importance to understanding the system. Convergence stability in distributed problems has been a topic of research for some time with the first work being performed by Vincent [31] for two designers, two design variables problems. Vincent introduced the game theory model for distributed design processes, which was investigated further by Lewis [39]. In Vincent's work each player alternates minimizing his or her local objective function value and communicates the associated design variables to the other player. Each step in this alternating process is a play in a sequential game. After repeated playing of the sequential game, the players either converge to a solution or diverge and continue playing indefinitely. When the players converge, they converge to a specific point called the Nash, or non-cooperative equilibrium [40].

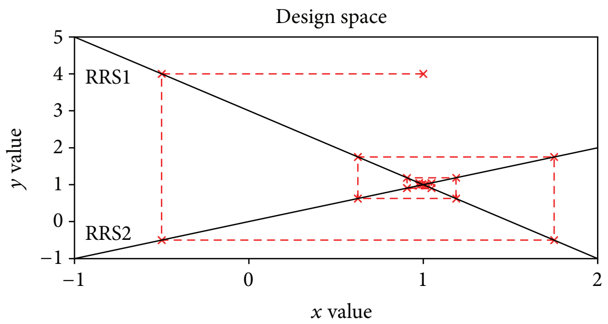


FIGURE 1: Two subsystems—Nash equilibrium.

Mathematically, in a two-player game a set of solutions described by the vector pair (x_1, x_2) are a Nash solution, (x_{1N}, x_{2N}), if they fulfill the requirements outlined in (1) as

$$\begin{aligned} F_1(x_{1N}, x_{2N}) &= \min_{x_1} F_1(x_1, x_{2N}), \\ F_2(x_{1N}, x_{2N}) &= \min_{x_2} F_2(x_{1N}, x_2). \end{aligned} \tag{1}$$

In (1), F_1 and F_2 are the objective functions for player 1 and player 2 who control design variables x_1 and x_2, respectively. A solution pair (x_1, x_2) that meets the criteria in (1) is a Nash solution because the pair is a minimum for both F_1 and F_2. Although in game theory the participants in a game are called players, in engineering design they are typically called designers or subsystems. In this work, the term subsystem is used more generally, reflecting the linear system basis of the work. The relationship demonstrated in (1) can be understood qualitatively as the point at which no subsystem can unilaterally improve his or her objective function [41]. This expression identifies Nash solutions through an optimization formulation, but they can also be expressed as the intersection of two sets defined by (2) as

$$\begin{aligned} (x_{1N}, x_{2N}) &\in X_{1N}(x_{2N}) \times X_{2N}(x_{1N}), \\ X_{1N}(x_2) &= \left\{x_{1N} \mid F_1(x_{1N}, x_2) = \min_{x_1} F_1(x_1, x_2)\right\}, \\ X_{2N}(x_1) &= \left\{x_{2N} \mid F_2(x_1, x_{2N}) = \min_{x_2} F_2(x_1, x_2)\right\}. \end{aligned} \tag{2}$$

The sets X_{1N} and X_{2N} are the rational reaction sets (RRS) or best response sets [42], which embody all the possible reactions or responses a subsystem may have towards a decision made by another subsystem. While determining the RRSs is not a trivial task, methods have been developed to approximate them for large systems [43]. One of these techniques is to calculate the RRS by taking the gradient of the subsystem's objective function with respect to its local design variables. This calculation is shown in (3) for the two designer problem studied by Vincent [31] as

$$\begin{aligned} F_1 &= x^2 + xy - 3x, \\ \frac{\delta F_1}{\delta x} &= 2x + y - 3 = 0, \\ F_2 &= 0.5y^2 - xy, \\ \frac{\delta F_2}{\delta y} &= y - x = 0. \end{aligned} \tag{3}$$

These RRSs are plotted with respect to the design variables, x and y, in Figure 1 along with an illustration of how the solution process converges to the Nash equilibrium at $(1, 1)$. We have included the dotted line in Figure 1 to demonstrate the order of the decisions made by the two subsystems. The subsystems iterate sequentially and begin with the initial conditions set to $(1, 4)$. The x's in Figure 1 shows the actual discrete decision occurring through the repeated decisions, identical to the plays of a sequential game. In Figure 1, subsystem 1 adjusts the value of x and subsystem 2 responds by adjusting y. For the decision making process in Figure 1, the design variable values converge to the Nash equilibrium, defined by the intersection of the players' RRSs. For two subsystem problems, Vincent defined stability criteria based on the subsystems' RRSs [31]. This work was extended by Chanron and Lewis to examine convergence more generally when there are more than two players controlling multiple design variables [3]. Convergence was shown to be a function of the relative slope of the designer's rational reaction sets and linear system theory was applied for large scale problems [3–6]. Work by Smith and Eppinger has demonstrated the same principle using a different set of fundamental principles [44].

In this work, distributed design processes are modeled using linear system theory, using the same approach developed by Chanron and Lewis. Similar to game theoretical models, each subsystem is assumed to solve its optimization problem at distinct and discrete instants in times that are identical for every iteration based on the process configuration. We assume that a state space model has already been developed for the distributed design systems analyzed in this paper. A detailed description of how to create these models can be found in [3]. The general formulation for a subsystem's objective function using nomenclature from linear system theory is shown in (4).

$$F_n = X^T AX + Y^T BY + X^T CY + DX + EY + F. \tag{4}$$

In this representation of the nth subsystem's quadratic objective function, F_n, X is a vector of length i which contains the i local design variables while Y is a vector of length j which contains j nonlocal design variables. The coefficients associated with the second-order elements of F_n for the local design variables are contained in the diagonal $i \times i$ matrix A while the coefficients associated with the non-local design variables are contained in the $j \times j$ matrix B. In this representation the A matrix is formulated as a diagonal to decouple the subsystem's local design variables from one another. This guarantees that each design variable value can be determined independently and a specific RRS can be formulated for each design variable. When these variables are coupled, the design system can still be represented using the form in (4). However, to do so, a change in variables must be performed to decouple the values from one another. The representation in (4) is examined in more depth in [3].

Although the local design variables must be decoupled, it is acceptable for the local and non-local design variables to be coupled together through the coefficients in the $i \times j$ C matrix. The remaining two vectors in (4) capture the linear elements of the system for the local and non-local design variables and have length i and j, respectively. The term F is simply a scalar and does not play a significant role when analyzing the system stability or transient response. The important elements in (4) emerge when the gradient is taken with respect to the local design variables. Setting this gradient equal to zero results in i decoupled equations that represent the subsystem's RRS. After the RRSs are found for each subsystem, there are m equations, where m is the total number of design variables controlled by subsystems. The RRS is shown in vector form in (5) as

$$\frac{\delta F_n}{\delta X} = 2AX + CY + D = 0. \tag{5}$$

Equation (5) specifies how this subsystems will respond and suggests the system's overall transient response is related to the matrices A, C, and D for each subsystem. Using these matrices, Chanron developed the discrete state-space-based representation to model the design systems using the update relationship and stability criteria in (6)–(9) as

$$X_s^{k+1} = \Phi X_s^k + \Gamma, \tag{6}$$

$$\Phi = -\frac{1}{2}\begin{bmatrix} A_1^{-1} & & & 0 \\ & A_2^{-1} & & \\ & & \ddots & \\ 0 & & & A_m^{-1} \end{bmatrix}\begin{bmatrix} 0 & C_{12}^T & \cdots & C_{1m}^T \\ C_{21}^T & 0 & & \vdots \\ \vdots & & \ddots & \\ C_{m1}^T & \cdots & & 0 \end{bmatrix}, \tag{7}$$

$$\Gamma = -\frac{1}{2}\begin{bmatrix} A_1^{-1}D_1^T \\ \vdots \\ A_m^{-1}D_m^T \end{bmatrix}. \tag{8}$$

In (6), the subscript s denotes that X_s^k is a vector of all the system design variables and the superscript denotes the iteration number which is consistent with linear system theory. Since (6) describes the relationship between the subsystems, X_s^{k+1} is length m containing all the design variables controlled by the subsystems. The design variable values at the $(k+1)$th iteration are a function of the previous design variables at the kth iteration; they are expressed as X_s^k multiplied by a matrix Φ plus a constant Γ. The derivations for Φ and Γ can be found in [3] and are summarized in (7) and (8). Equation (6) was generalized in [5] to be applicable to scenarios where decisions are made asynchronously.

The matrix Φ captures design variable interactions between quadratic elements found in the A and C matrices while the vector Γ captures interactions between quadratic and linear elements found in the A and D matrices, respectively. To populate Φ and Γ, the appropriate A, C, and D matrices must be used and can be determined by examining which subsystem controls the design variable associated with the row being populated. The resulting dimensions for Φ and Γ are $m \times m$ and $m \times 1$, respectively. When examining system stability, only the Φ matrix needs to be considered, and if the condition described by (9) is met then the system is stable

$$r_\sigma(\Phi) < 1. \tag{9}$$

In (9), $r_\sigma(\Phi)$ is the spectral radius, or magnitude of the largest eigenvalue of the matrix Φ [45]. The relationship in (9) specifies that for stable systems, Φ must have a spectral radius less than 1. This is the same stability criteria used for the closed loop state space representations of discrete control systems [46]. In addition to the examination of linear system stability, a case for nonlinear RRSs has also been investigated in [47] using similar criteria. Another extension of this convergence work was performed by Gurnani and Lewis who demonstrated that the introduction of "mistakes" into the design process could cause some inherently unstable problems to converge to a solution [48].

Analyzing system stability is the first step to characterizing transient response. In this paper, we examine configurations that (9) identifies as stable and develop an approach to differentiate between them based on their convergence time. In the next section, examining the transient response of distributed design systems is discussed and the idea of solution process architecture is introduced as a key factor determining the transient response of a distributed design system.

2.4. Transient Response of Distributed Design Systems. The transient response of a distributed design process has two principle aspects. The first is the shape of the transient response. This shape depends on the eigenvalues of the system being analyzed and could be sinusoidal, exponential, or some combination of responses. An example transient response shape is shown in Figure 2, which plots the value of design variables x and y from (3).

Since design decisions occur at a specific instant in time, the design variable plots in Figure 2 are staircase plots representing discrete design variable values. Figure 2(a) tracks the value of x while Figure 2(b) tracks y. In this case, both variables exhibit a decaying sinusoidal response as they approach their equilibrium value at $(1, 1)$ from a starting location of $(1, 2)$. Identifying the shape of the convergence curve for a distributed design process is an important area of future research, but this work focuses on examining the second aspect of transient response, convergence time.

The convergence time in this work is measured by the number of iterations required for the subsystem to reach an equilibrium solution. We use iterations to evaluate convergence time because they are a dimensionless characterization of the system that can be easily translated to the time domain by either mapping estimated task times directly or by leveraging the work transformation matrix approach used in [44]. Although a significant amount of rigor has been brought to the analysis of the stability characteristics of distributed design processes, the convergence time of these processes or of MDO processes in general has focused more on practical implementation than investigating controlling features. Techniques like the critical path method [49] and project evaluation and review technique [50] are the foundational approaches used in network-based project planning. The techniques for network-based project planning provide approaches to organize and execute the design tasks inherent to MDO processes. In the context of distributed design processes, there are tools to specify the ordering or

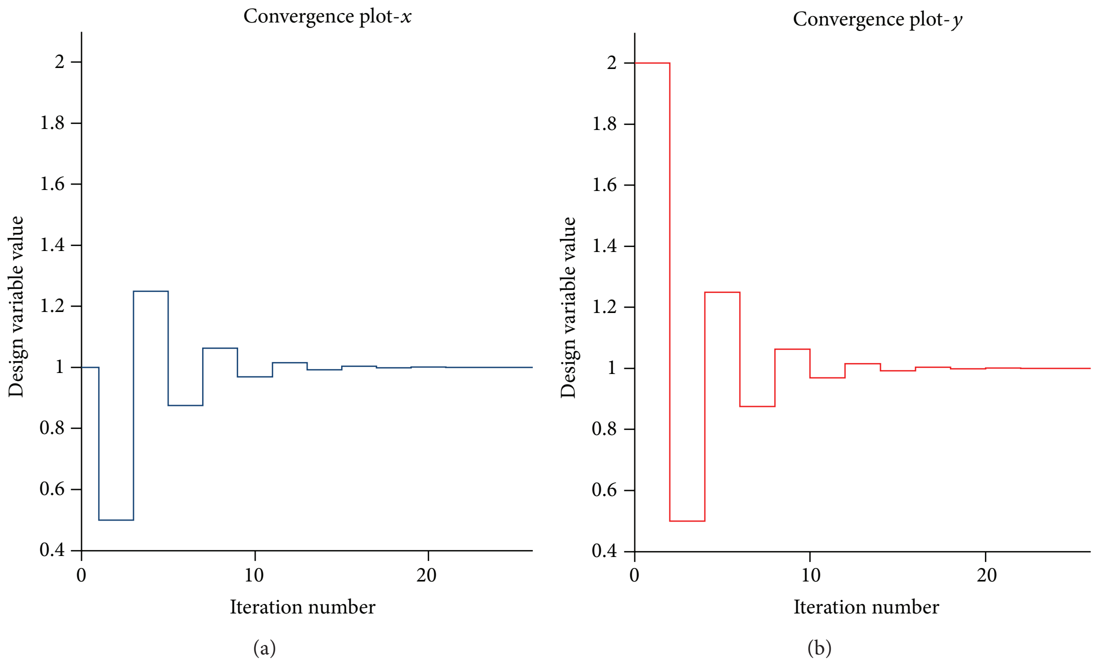


Figure 2: Two design variable convergence plot.

organization of the process which is discussed in Section 2.5. These techniques have been refined to introduce technology like Monte Carlo simulation in [51] to account for random task duration, graph theory in [52] to enable probabilistic organization, and feedback and precedence relationships in [53] to account for required work flow.

More recent formulations have leveraged advances in computing power to prescribe the fastest converging process organization. One of these techniques is an extension of the Design Managers Aid for Intelligent Decomposition (DeMAID) method and is used to reduce the time required for designers to converge to a final solution. In this extension, Rogers utilizes the global sensitivity equations [54] with a weighting scheme to predict an optimal ordering of designers in iterative design loops [55]. The approach taken by Rogers succeeded in reducing the overall design time required for iterative loops in DeMAID when designers are ordered sequentially. This ordering was partially based on an analysis of design structure matrices (DSMs), which were developed in [56] as a means to organize and visualize the coupling between design tasks.

DSMs were also used to minimize the number of feedback loops for sequential processes in [57]. An approach using DSMs to represent the probability of task repetition and durations with Markov chains was presented in [58] and a case was made for estimating the stability and convergence rate of concurrent tasks in [44]. In this work, a DSM-based transformation matrix was used to link design tasks based on the amount of rework tasks generated for one another. An eigenvalue analysis was used to determine the strength of these links and the task coupling. While this approach examines the basic mechanics of the distributed system, it does not account for changes in the ordering of design systems and does not provide sensible bounding conditions for the amount of time required for the system to converge. Although these techniques suggest orderings for the design process, only [44] provides a prediction for the overall convergence time. Simulation-based techniques have predicted the convergence time for concurrent engineering in [59], for overlapping tasks in [60] and for using DSMs in [47] which was further refined in [36].

A general convergence model for use in specifying architectures was presented in [61]. Simulation has also been used to tie process architecture to solution quality and suggest strategies to realize better products in [62]. Another approach attributed design process delays to incomplete sharing of design information and provides a dynamic work transformation model to determine when incomplete sharing occurs in [63]. A case was made for ordering design tasks to reduce the amount of uncertainty inherent to the problem in [64]. An analytical technique derived specifically for distributed design processes examined the relationship between the system transient response and the ordering of the solution process in [65] for two-subsystem systems.

This work differentiates itself from other techniques by providing an approach to determine the upper bound for the number of iterations required for a distributed design process to reach equilibrium. It does not require simulation to evaluate a proposed architecture and provides estimates based on the coupling of the system's component subsystems. This provides a computationally inexpensive initial evaluation of process architectures where the most promising architectures can be evaluated using more expensive, time consuming techniques. It differentiates itself from [65] through its applicability to large design systems and from previous work by requiring no system simulation. Before

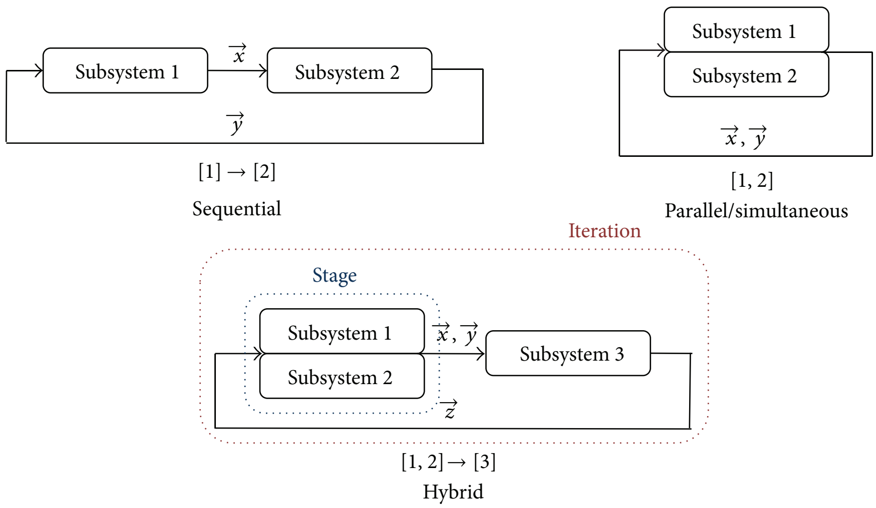


Figure 3: Potential process architectures.

presenting the approach, however, the concept of solution process architecture is introduced and its influence on the systems transient response is discussed.

2.5. Solution Process Architecture. The solution process architecture is the organization or structure used to solve a distributed design process. This structure can include both sequential and simultaneous solution processes. Pure sequential or pure simultaneous process architectures are the two extreme cases for process architectures. In Figure 3, a simple diagram illustrates the iterative process for a purely sequential architecture, a purely simultaneous architecture, and a hybrid approach which utilizes both sequential and parallel elements.

Each of the three process architectures described in Figure 3 represents a single iteration of the solution process. Repeated iterations of the architecture are used to solve a specific design process. These iterations can be further broken down and a single subsystem or a set of subsystems arranged in parallel with one another is called a stage. The difference between iteration, stage, and subsystem is shown for the hybrid architecture in Figure 3. The number of stages in process architectures depends on the number of subsystems and the process architecture chosen. For purely sequential process architectures, the number of process stages is always equal to the total number of subsystems. In contrast simultaneous, or parallel, process architectures always consist of a single stage. The number of process stages for sequential and parallel process architectures provide an upper and lower bound, respectively, for the number of stages in hybrid process architectures. For example, the hybrid process architecture in Figure 3 has two stages.

The stability criteria developed by Chanron and Lewis and shown in (7) is applicable to the parallel process architecture in Figure 3. In a recent extension, this criteria was refined to encompass simultaneous and hybrid systems as well [5]. This extension represents the design system in the same form as (6) but makes some allowances for process architecture changes. It was also demonstrated that process architecture

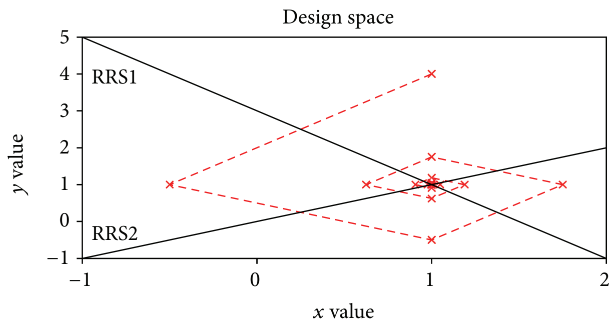


Figure 4: Two subsystem—parallel design architecture.

has a significant impact on both the system stability, and the transient response. The location of the equilibrium, however, remains unchanged.

Some approaches have also utilized process architectures with partial overlapping between subsystems, where a new subsystem begins solving its optimization problem after another has already begun but not finished its optimization [60]. Since this is an initial investigation, we consider the three architectures shown in Figure 3 and assume that all subsystems begin a stage simultaneously with no overlapping.

To illustrate the relationship between process architecture and convergence time, the system described in (3) is simulated using a sequential process architecture in Figure 1. Since the equilibrium design variable values are known a priori, the process is said to have converged if all the design variables values are within 2% of their final values. Given this criterion, convergence takes 26 iterations for a sequential architecture. In contrast, the convergence plot for a parallel architecture for the two subsystem problem is shown in Figure 4.

A comparison of the dotted line representing the convergence path in Figures 1 and 4 demonstrates very different paths from the same starting location at $(1, 4)$ to the same Nash equilibrium at $(1, 1)$. This difference is caused by the way the designers share design variables. The difference in the

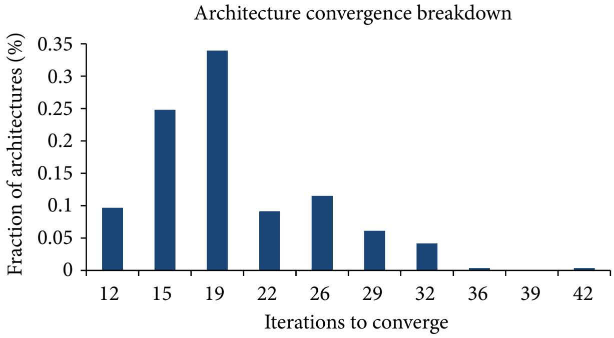


Figure 5: Convergence times for the five designer problem.

path taken by the designers also translates to a difference in solution time, with the parallel system requiring only 22 iterations to converge using the same criteria as used in Figure 1. A more comprehensive examination of the differences between these two architectures has been summarized in a study of convergence time in distributed design systems found in [65].

In a two-designer system, there are only two potential design architectures. However, as the number of designers increases, there are a large number of potential design architectures that fall into the category of hybrid. These hybrid architectures can have a significant impact on the system stability and convergence time. For larger systems, there are a wide range of architecture options with very different associated convergence times. This can be demonstrated by considering the same problem used in [3] to study the stability of large systems. This problem is an unconstrained five-designer problem with sixteen unique design variables. The convergence times for the different randomly generated architectures simulated from a variety of initial starting locations are plotted in Figure 5.

In Figure 5 the architectures are grouped into bins based on the number of iterations required to converge, demonstrating the wide range of convergence times that can be obtained by changing the process architecture. The height of the bars indicates the number of process architectures with a specified range of convergence times, shown on the x-axis. The mean convergence time for the simulated architectures is 25 iterations with the fastest convergence of 14 iterations and the slowest convergence of 42 iterations.

Although all of the process architectures studied in Figure 5 have a stable Nash equilibrium, there are some cases when changing the process architecture can change the system stability. This is because changing the process architecture also changes the system eigenvalues as shown in [5]. The influence of architecture, called topology, on convergence, stability, and equilibrium is also studied for real systems in [66, 67]. Where Braha and Bar-Yarn develop a descriptive approach to characterize decision networks in their work, this paper analyzes normative models to characterize convergence rate. Since system stability can be assessed by analyzing the system eigenvalues, it is proposed in this work that the eigenvalues associated with process architectures can be used to evaluate those systems' transient response. Examining the relationship between the system eigenvalues and the transient response is the focus of Section 3.

Table 1: Spectral radius experiment parameters.

Parameter	Value
Number of designers	4 to 10
Number of design variables	4 to 15
A	−20 to 20
C	−20 to 20

3. Results and Discussion

Since an eigenvalue analysis is used to determine the stability of specific solution process architectures, it is natural to examine eigenvalues to determine the architectures' convergence times. Existing linear systems theory evaluates system eigenvalues to determine settling time, natural frequency, modal response, system damping, and a number of additional properties. Furthermore, empirical evidence in Figure 6 suggests a relationship exists between the spectral radius of a distributed design system and the convergence rate.

The data in Figure 6 was generated by evaluating the spectral radii associated with ten different solution process architectures for five randomly created distributed design systems. The data used to create these systems is shown in Table 1 and the systems themselves are in the form of (7) and (8). In order to reduce the number of possible parameters in the experiment that may bias the result, the values in the D vector were set to zero to guarantee all the design systems had equilibrium solutions at the origin. Also, each subsystem was given local control of one design variable while the remaining variables were randomly allocated to the different subsystems.

The process architectures with spectral radii greater than 1 were not included in Figure 6 because they had unstable equilibrium solutions. To minimize the impact of starting location on the convergence behavior, each data point in Figure 6 is the average of twenty simulations started from a set of different points. Similar to the previous simulation, the process is defined to converge when the design variables are all within 2% of their final values. Although there is some correlation between the spectral radius and the mean number of iterations to converge, the circled architectures demonstrate that this mapping is not monotonic as some previous work has suggested [3]. Systems with the same spectral radius can have very different convergence times. This variation is demonstrated less dramatically across several of the other architectures with smaller spectral radii as well. Even when design systems have the same convergence time, the spectral radii of those systems can vary significantly. The systems generated in this section are used to experimentally support the approach outlined in this paper.

As demonstrated in Figure 6, the spectral radius is insufficient to quantify the convergence time of a distributed design process. For linear control systems, the real and imaginary components of the eigenvalues are used to determine the

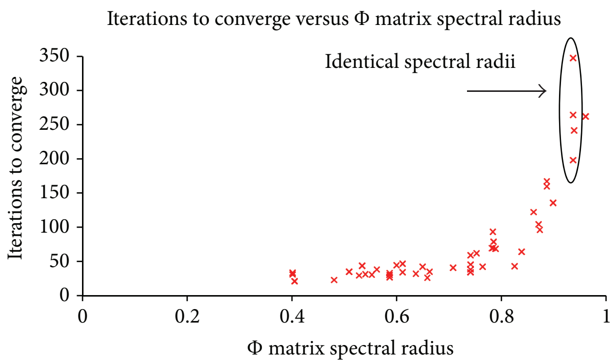


FIGURE 6: Iterations to converge versus phi spectral radius.

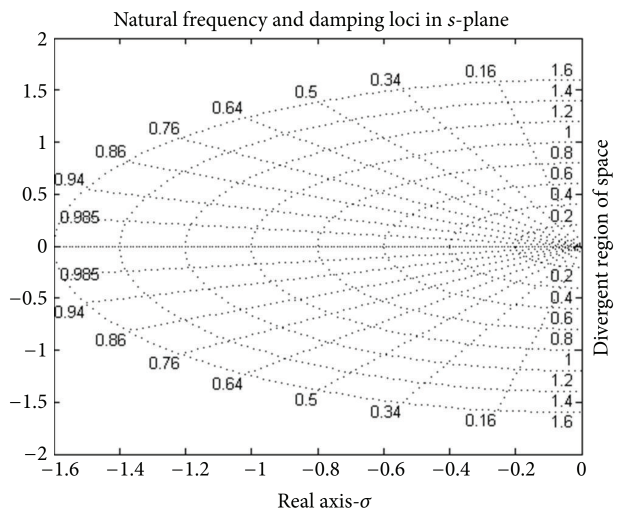


FIGURE 7: s-Plane—natural frequency and damping loci.

natural frequency and damping ratio of the system. Using these quantities, an upper bound can be determined for the overall system convergence time. This approach is adapted in this work to quantify convergence time in decentralized design systems. However, determining the natural frequency and damping ratio for distributed design processes is challenging and is the focus of the remainder of this paper.

Since decision networks are inherently discrete systems, they are modeled as discrete time state space systems. To determine the natural frequency and damping ratio of general discrete time systems, they are converted to continuous time approximations. However, distributed design processes are unique because they do not possess any underlying continuity. A continuous time model is, therefore, a true abstraction of the actual behavior.

In this paper, two techniques are examined to convert between the continuous and discrete time domain. These techniques are the zero-order hold and the bilinear, or Tustin, approximation. These two techniques are chosen because they are commonly used in the systems literature and the Tustin approximation is generally acknowledged as the preferred technique for this type of conversion [46]. This paper is not an exhaustive analysis of techniques that can be used to transform discrete distributed design systems into continuous time representations. Instead, it examines two prevalent approximations and identifies the one which results in a continuous time model that can exactly reproduce subsystem decisions at the appropriate discrete point in time.

Before any of these analyses can be conducted, however, a discrete time state space model must first be created to accurately represent the system. Creating these state space models is the topic of [3, 5] and in this paper all state space models are created using these processes. The fundamental difference between discrete time and continuous time state space models is that their eigenvalues are plotted in the z-plane rather than the s-plane. The transforms required to plot these eigenvalues are discussed in the following section.

3.1. Laplace and Z Transforms for Distributed Design Models. The primary challenge in quantifying the convergence time of distributed design processes is successfully transforming them to the continuous time domain. The advantage of a continuous time representation of a system is that it enables the eigenvalues to be plotted in the s-plane to capture the frequency response characteristics of the system. However, to plot a system in the s-plane, it must first be represented as a set of algebraic equations in a single value, typically s, by taking the system's Laplace transform.

The roots, poles, or eigenvalues of these algebraic equations have both real and imaginary components and are plotted in the complex s-plane where the real components are plotted on the x-axis while the imaginary components are plotted on the y-axis. The basic information provided in this plot is a system's natural frequencies, damping ratios, and stability characteristics. Furthermore, these properties can often be determined through inspection. A representation of the s-plane generated using the *sgrid()* command in MATLAB which includes lines of constant natural frequency and damping is shown in Figure 7.

In Figure 7, the lines radiating outward from the origin into the second and third quadrants are lines of constant damping ratio, ζ. The concentric arcs in Figure 7 centered at the origin and extending through the second and third quadrant are curves of constant damped natural frequency, ω_n. Systems with eigenvalues located on the left hand side of the y-axis are stable and settle to an equilibrium value in finite time, while systems with eigenvalues located on the right hand side of the y-axis are unstable and diverge. If the eigenvalues are on the y-axis in the s-plane, they are saddle points and the system oscillates forever, without moving closer or further away from the equilibrium value.

In the same way, an s-plane representation captures the characteristics of a continuous time system, and the z-plane can be used to capture characteristics of discrete time systems. To represent a system in the z-plane, the z transform of the system's time-invariant difference equations is taken to create an analytical expression in terms of a single variable, z. Once again the roots of this expression are plotted, this time in the z-plane. Although the s-plane and z-plane are both complex planes, the z-plane's properties are significantly different. A plot showing contours with constant natural frequency and

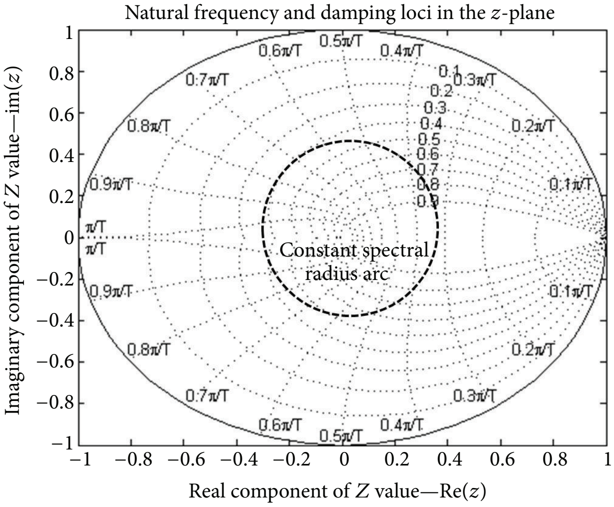


Figure 8: z-Plane—natural frequency and damping loci.

damping in the z-plane was generated using the MATLAB command *zgrid()* and is shown in Figure 8.

The stable region of the z-plane is shown in Figure 8 and corresponds to the region circumscribed by the unit circle, which graphically shows the stability criteria in (9). The region outside the unit circle is unstable and the unit circle itself is the set of saddle points. The region inside the unit circle has two sets of contours. The contours originating from the x-axis at $(1, 0)$ are curves of constant damping ratio. The other set of contours, perpendicular to the unit circle, are arcs of constant natural frequency.

Examination of the z-plane explains why eigenvalues, which correspond to roots of the z transform, with the same magnitude have a wide range of convergence times in Figure 6. Although these eigenvalues have the same magnitude, they map to very different natural frequencies and damping ratios. The damping ratio and natural frequency of a system can be used to determine the system's convergence time. For a second order, linear time invariant system, the convergence time, called the settling time, can be related to the damping ratio and natural frequency using (10) [68] as

$$t_s < 4.6\tau,$$
$$\tau = \frac{1}{\omega_n \zeta}. \tag{10}$$

In (10), t_s is the settling time, which is defined to be the time required for a system to converge to within 2% of its final value as measured from its initial value. The variable τ is the system's time constant which is the inverse of the natural frequency, ω_n, multiplied by the damping ratio, ζ. In this paper, we examine the applicability of (10) to systems that are not inherently continuous and use it to create an upper bound for their convergence time.

To apply (10), the natural frequency and damping ratio associated with the system roots is required. Unlike an s-plane representation, this value cannot be read directly from a plot in the z-plane and it depends on the sampling period for the discrete system. This sampling period is also critical to map system roots between the z-plane and the s-plane, which is a desirable transformation since the analysis of system transient response is often conducted in the frequency domain represented by the s-plane. The relationship in (10) is an example of an approach to approximate the convergence time of a continuous time system. The relationship between the complex variables in the z- and s-planes is mathematically expressed in (11) [46] as

$$z = e^{Ts}. \tag{11}$$

In (11), z is a complex number representing a root in the z-plane while s is the associated root in the s-plane. The value T is the sampling period for the system, usually measured in seconds. Since T can be any positive value, a single point in the z-plane can be mapped to many values in the s-plane depending on the sampling rate. The entire left hand side of the s-plane is represented by the unit circle in the z-plane because both encompass the entirety of the stable regions for a linear system.

The relationship expressed in (11) maps points between the z- and s-planes and can be applied to points in discrete time systems sampled at the appropriate rate. It is leveraged later in this work to transform eigenvalues of distributed design systems. However, to this point it has not been used to map poles of distributed design processes from the z- to s-plane because T has remained indeterminate. To facilitate this mapping, an analogy for the sampling rate of distributed design processes is discussed in the next section and the zero-order hold and bilinear transformation are presented to transform discrete systems into continuous systems and determine T.

3.2. Distributed Design Sampling Rates and Transformations. Distributed design systems are unique because they are truly discrete systems. A distributed design process is composed of a set of point discontinuities that represent specific decisions made at an instant in time. Since the process is governed by discrete iterations, there is no underlying continuity between decision points. This contrasts with most other systems that are fundamentally continuous, but sampled in time to create a discrete representation. The sampling rate for these continuous time systems is of critical importance, because low sample rates can inhibit the accurate reproduction of the continuous time signal. Furthermore, relationships used to analyze and reconstruct the signal for discrete systems generally require a specific sampling rate. To capture the system's transient response for this reconstruction, it must be sampled at a minimum rate called the Nyquist frequency [68].

When a continuous time system is sampled at a rate above the Nyquist frequency, the resulting set of points is an abstraction of the actual continuous time system. Since distributed design processes are by their nature discrete, creating a signal in the continuous time domain is not a reconstruction of the signal but an abstraction of the system's true behavior. This distinction restricts which linear system tools and techniques can be used to analyze and model distributed design processes. In this work, the zero-order

Table 2: Eigenvalues associated with system approximation.

Approximated eigenvalues	
Zero-order hold method	Tustin approximation
−0.367	−0.363
$-0.482 \pm 1.33i$	$-0.740 \pm 1.43i$
−1.62	−1.34

hold and the Tustin, or bilinear, approximation are used to construct a continuous time representation for distributed design processes to determine T. A simulated annealing algorithm was used to determine the sampling rates that led to the most accurate overall representation. The objective function for the simulated annealing algorithm was the distance between the discrete time points and the continuous time curve as measured using an L_2 norm. The design variable in the optimization formulation shown in (12) is T.

$$\text{Minimize: } f(T) = \sqrt{\sum_{j=1}^{k}\sum_{i=1}^{m}\left(x_i^j - g_i^j(T)\right)^2} \tag{12}$$

$$\text{Subject to: } T > 0.$$

In (12), x_i^k is the design variable value for the ith design variable at the kth iteration. The term $g_i^j(k)$ is the ith design variable value as determined from the continuous time model generated using either the zero-order hold or Tustin approximation at the jth iteration. Finally, T, the sampling rate, is the design variable value for the optimization and is used to develop the continuous time expression $g_i^j(k)$. From this experiment, it is found that a sample rate of 1 sample per second most accurately represented the discrete time system in continuous time for both the Tustin approximation and zero-order hold.

Even though both transformations used the same sampling rate, the zero-order hold and Tustin approximation did not construct identical continuous time representations. As an example, the four largest eigenvalues for one of the simulated systems are summarized in Table 2 for both the zero-order hold representation and the Tustin approximation.

Both approximations result in models with different continuous time eigenvalues. Examination of the eigenvalues in Table 2 shows that both methods identify almost the same first eigenvalue for the design system. However, the second eigenvalue identified is significantly different for each method.

When determining the appropriate sampling rate, the zero-order hold representations are more accurate when considering the distance between the continuous and discrete time points as measured using the objective function in (12). The difference between these two representations is more obvious when plotted. The curves generated for the first design variable of the system summarized in Table 2 are plotted in Figures 9 and 10, respectively.

In Figures 9 and 10, time is plotted on the x-axis and the systems reach their equilibrium value after eleven

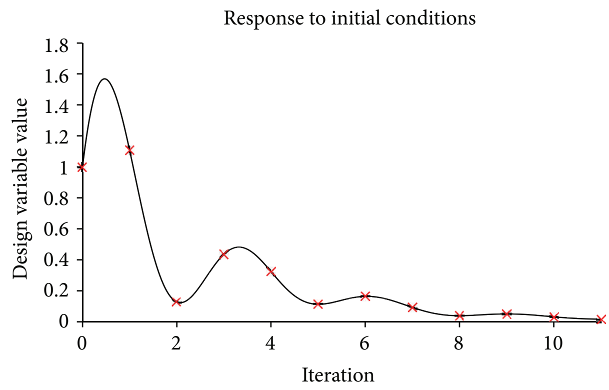


Figure 9: Zero-order hold method.

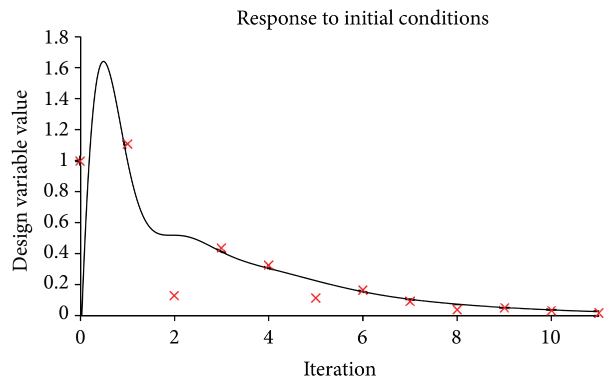


Figure 10: Tustin Approximation.

iterations. The design variable value is plotted on the y-axis, starting from an initial value of one and converging to a value of 0. Both plots show the first design variable for the simulated system and are created using the *initial()* function in MATLAB. The actual discrete design variables, as determined through simulation, are marked by x's in both Figures 9 and 10.

The curve in Figures 9 and 10 is the continuous time approximation for the discrete system. Inspection of the two figures shows that the approximation based on the zero-order hold passes through every discrete design variable value. This representation is more desirable because each discrete design variable represents an actual decision in the design process. The Tustin approximation does not pass through every design point in Figure 10 and does not accurately reproduce the design process.

The zero-order hold produces an accurate continuous time model for the system because its assumptions match extremely well to the fundamental mechanics of distributed design processes. The zero-order hold converts a discrete time signal to continuous time by holding a single sample's values constant over the sampling period. This signal reconstruction technique mirrors distributed design processes, where each subsystem assumes constant non-local design variable values when solving their local optimization problem for a single decision step in the process.

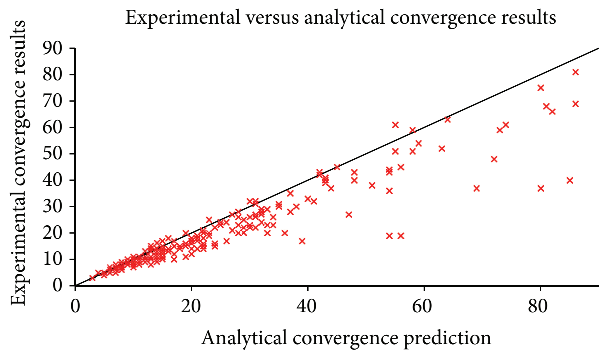


FIGURE 11: Experimental versus analytical convergence results.

Using the zero-order hold to transform a discrete state space model of a distributed system enables the analysis of the system using classical control techniques. The identification of the appropriate sample rate for these models makes the transformation from the z-plane to the s-plane simpler as well. In the next section, an analytical relationship based on these findings is introduced to transform discrete time eigenvalues to their continuous time equivalents. These continuous time equivalents are then analyzed using control theoretic relationships and the results are validated experimentally.

3.3. Analyzing the Convergence Time of Distributed Design Processes. In this section, the concept of the system time constant is examined for distributed design processes and its applicability to discrete time systems is demonstrated empirically. The mathematical relationship for continuous time systems between the natural frequency, damping ratio, and time constant is summarized in (10). It is demonstrated using the zero-order hold that it is possible to model distributed design processes as continuous time systems, and it is proposed that the relationship in (10) can be used to create an upper bound on the convergence time of distributed design processes as well. Since (10) requires only information about the damping ratio and natural frequency, a complete continuous time model of the system is not required. Instead, a simpler relationship can be used to determine these properties from the discrete time eigenvalues. These relationships are provided in (13) and (14) [46] as follows:

$$\frac{\zeta}{\sqrt{1-\zeta}} = \ln\left(\frac{|z|}{\angle z}\right), \tag{13}$$

$$\omega_n = -\ln\left(\frac{|z|}{T\zeta}\right). \tag{14}$$

In (13) and (14), ζ is the damping ratio, and ω_n is the natural frequency. To determine $|z|$, and $\angle z$ it is assumed that the z-plane form of the eigenvalues is $z = a + bi$, with a being the real part of z and b being the imaginary part. Using a and b the magnitude, $|z|$ is the vector sum of a and b while $\angle z$ is the $\arctan(b/a)$. The parameter T in (14) is the sampling period for the system. Recall that this does not have a direct analogy for distributed design systems but is discussed in Section 3.2 and is 1 sample per iteration.

Equations (10), (13), and (14) are used to evaluate the convergence time for 250 distributed design systems using the parameters outlined in Table 1. The settling time for each system is analytically determined and plotted in Figure 11 against the actual settling time for the system as determined by simulation.

The line in Figure 11 is shown as a reference to compare the convergence time determined by simulation to the analytical convergence rate predictions. It has a slope of one to differentiate between cases where the approach over and under predicted the system convergence time. For almost all the systems, the analytical results, determined using the settling criterion from (10), provide an upper bound for the convergence time of the design systems. Approximately 7.6% of the systems are above the solid line in Figure 11, which means that the approach underestimated the convergence time. The furthest system above the line exceeded the predicted maximum convergence time by 6 iterations, which is 10% of its convergence time. All other points, however, are at most 2 design iterations greater than their predicted value which means that the prediction was very close to the simulated convergence time.

One advantage to this approach is its efficiency, since it only considered the largest pole of the system, required no simulation, and is independent of the starting location. The dynamic behavior of many of the systems may have been understated with this simplification. For example, the system whose convergence was underpredicted by 6 iterations was composed of 6 unique subsystems collectively controlling 32 unique design variables. To capture the dynamics of more sophisticated design problems using a single eigenvalue may be insufficient to appropriately model the system.

In spite of only considering one eigenvalue, the approximation provides an upper bound for many of the systems. One reason for this may be the discrete nature of distributed design processes. A continuous time model for a truly discrete system overestimates the settling time because the gaps between discrete time points enable it to skip past the peaks of the continuous time curve. These peaks may remain outside of the 2% settling time longer than the discrete points unless a discrete time point is located on a peak. Additional areas of future work to increase the effectiveness of this approach for large systems are discussed in the next section.

4. Conclusions

In this paper, the transient response of distributed design processes, as modeled as a two-player continuous game, is characterized. An eigenvalue analysis formed the basis for this examination and concepts native to continuous control theory are used to evaluate and approximate the transient response of distributed design systems. The transient response is broken down into two components: shape and convergence rate. This paper focuses on the convergence rate of distributed design processes and it is shown that the convergence rate cannot be assessed using only the magnitude of the system's largest eigenvalue.

Instead of examining eigenvalue magnitudes, this paper modeled inherently discrete distributed design processes as continuous time systems. Two commonly used approximations, the zero-order hold and Tustin approximation, are evaluated to determine the parameters required to analyze distributed design processes in continuous time. Both approximations suggest that the best sampling period, a key parameter in the conversion between discrete and continuous time systems, is 1 sample per iteration. It is also shown that the Tustin approximation does not provide an accurate reproduction of distributed design processes in continuous time. On the other hand, the zero-order hold provides a good approximation for these systems because its assumptions match well with the fundamental mechanics of distributed design processes.

By using a continuous time approximation, the convergence rate aspect of a transient response is evaluated. This analysis uses a second-order approximation for the system. It is demonstrated that for many systems a second-order approximation provides a reliable upper bound for the number of iterations required for a design system to converge to an equilibrium solution. In cases where the approach is unable to provide an accurate estimate of the system transient response, the prediction remained in the neighborhood of the time required for the system to converge as determined by simulation.

One of the major contributions of the approach presented is that it enables the evaluation of different design process architectures without the need to simulate the distributed design system. Another contribution of this approach is that it validates an extension of the linear system theory analogy used to model distributed design systems. It identifies the role of sampling in the overall system response and emphasizes the truly discrete nature of distributed design problems. For systems with dominant closed loop poles the approach provides a mathematically provable upper bound for the system convergence time. Even when this criterion is not met, the approach is able to bound the convergence time for most cases without the computational costs associated with simulation. This provides a filter to identify potential process architectures that are worth committing additional resources to investigating. Finally, it provides a basis for the analysis of increasingly complex distributed design systems by examining the basic challenges to predicting system transient response.

Future work will focus on extending the approach to analyze large systems by considering more than one eigenvalue. Performing this analysis will require the identification of the most influential closed loop poles to create a higher-order model of the system. This model will provide greater fidelity and predictive ability while reducing the computational costs to simulate the system. Another area of future work is in the application of model reduction techniques to reduce the size of the models for distributed design systems. Currently distributed design models require a number of states equal to the number of shared design variables. Reducing the model will minimize the computational cost to analyze these systems both with respect to their stability and transient response characteristics. Incorporating our technique with a comprehensive simulation-based approach to evaluate process architecture is another area of future work. Combining the fidelity of simulation-based techniques with the initial response characterization from this work will reduce the computational burden to identify preferred process architectures.

Finally, the expanded applicability of control theory techniques demonstrated in this work provides a foundation to investigate principles used to analyze nonlinear systems. Translating these principles into techniques that provide meaningful evaluations of convergence times will enable the analysis of a broader class of distributed design problems.

Acknowledgments

The authors would like to thank the National Science Foundation, Grant DMII-0322783, and the Moog Corporation for their support of this work. Approved for Public Release; Distribution Unlimited: 88ABW-2013-2106, dated May 1, 2013.

References

[1] V. Chanron and K. Lewis, "A study of convergence in decentralized design," in *Proceedings of the ASME Design Engineering Technical Conferences and Computers and Information in Engineering Conference (DETC '03)*, 48782, pp. 765–774, Chicago, Ill, USA, September 2003.

[2] V. Chanron and K. Lewis, "A study of convergence in decentralized design processes," *Research in Engineering Design*, vol. 16, no. 3, pp. 133–145, 2005.

[3] V. Chanron and K. Lewis, "Convergence and stability in distributed design of large systems," in *Proceedings of the ASME Design Engineering Technical Conferences and Computers and Information in Engineering Conference (DETC '04)*, 57344, pp. 593–603, Montreal, Canada, October 2004.

[4] E. Devendorf, P. Cormier, and K. Lewis, "Development of a distributed design toolkit for analyzing process architectures," in *Proceedings of the 13th AIAA ISSMO Multidisciplinary Analysis and Optimization Conference (AIAA '10)*, 9029, Fort Worth, Tex, USA, 2010.

[5] E. Devendorf and K. Lewis, "The impact of process architecture on equilibrium stability in distributed design," *Journal of Mechanical Design*, vol. 133, no. 10, pp. 101001–101013, 2011.

[6] E. Devendorf and K. Lewis, "Quantifying the convergence time of distributed design processes," in *Proceedings of the ASME Design Engineering Technical Conferences & Computers and Information in Engineering Conference (DETC '11)*, 48377, Washington, DC, 2011.

[7] Y. C. Kim, L. H. Keel, and S. P. Bhattacharyya, "Transient response control via characteristic ratio assignment," *Institute of Electrical and Electronics Engineers*, vol. 48, no. 12, pp. 2238–2244, 2003.

[8] C.-L. Li, "Characterization of the equilibrium strategy of fuzzy bimatrix games based on *L-R* fuzzy variables," *Journal of Applied Mathematics*, vol. 2012, Article ID 824790, 15 pages, 2012.

[9] V. Vidyottama, S. Chandra, and C. R. Bector, "Bi-matrix games with fuzzy goals and fuzzy pay-offs," *Fuzzy Optimization and Decision Making*, vol. 3, no. 4, pp. 327–344, 2004.

[10] T. Maeda, "On characterization of equilibrium strategy of two-person zero-sum games with fuzzy payoffs," *Fuzzy Sets and Systems*, vol. 139, no. 2, pp. 283–296, 2003.

[11] S. Chien and A. Sinclair, "Convergence to approximate Nash equilibria in congestion games," *Games and Economic Behavior*, vol. 71, no. 2, pp. 315–327, 2011.

[12] I. Caragiannis, A. Fanelli, N. Gravin, and A. Skopalik, "Computing approximate pure Nash equilibria in congestion games," *ACM SIGecom Exchanges*, vol. 11, no. 1, pp. 26–29, 2012.

[13] L. I. de Castro, "Equilibrium existence and approximation of regular discontinuous games," *Economic Theory*, vol. 48, no. 1, pp. 67–85, 2011.

[14] V. M. Perez, J. E. Renaud, and L. T. Watson, "Reduced sampling for construction of quadratic response surface approximations using adaptive experimental design," in *Proceedings of the 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference (AIAA '02)*, 1587, Denver, Colo, USA, 2002.

[15] J. Sobieszczanski-Sobieski and R. T. Haftka, "Multidisciplinary aerospace design optimization: survey of recent developments," *Structural and Multidisciplinary Optimization*, vol. 14, no. 1, pp. 1–23, 1997.

[16] A. Khajavirad and J. J. Michalek, "A deterministic lagrangian-based global optimization approach for quasiseparable nonconvex mixed-integer nonlinear programs," *Journal of Mechanical Design*, vol. 131, no. 5, pp. 051009-1–051009-8, 2009.

[17] E. J. Cramer, J. E. Dennis, Jr., P. D. Frank, R. M. Lewis, and G. R. Shubin, "Problem formulation for multidisciplinary optimization," *SIAM Journal on Optimization*, vol. 4, no. 4, pp. 754–776, 1994.

[18] N. P. Suh, *Axiomatic Design: Advances and Applications*, Oxford University Press, New York, NY, USA, 2001.

[19] G. Pahl and W. Beitz, *Engineering Design: A Systematic Approach*, Springer, New York, NY, USA.

[20] S. Pugh, *Creating Innovative Products Using Total Design*, Addison-Wesley, Reading, Mass, 1996.

[21] M. M. Wiecek and J. A. Reneke, "Complex systems decomposition under uncertainty and risk," in *Proceedings of the 6th World Congress of Structural and Multidisciplinary Optimization*, Rio de Janeiro, Brazil, 2005.

[22] H. M. Min, N. F. Michelena, P. Y. Papalambros, and T. Jiang, "Target cascading in optimal system design," *Journal of Mechanical Design*, vol. 125, no. 3, pp. 474–480, 2003.

[23] B. Wujek, J. E. Renaud, and S. Batill, "A concurrent engineering approach for multidisciplinary design in a distributed computing environment," in *Multidisciplinary Design Optimization: State of the Art*, N. Alexandrov and M. Y. Hussaini, Eds., pp. 189–208, NASA, Hampton, Va, USA, 2003.

[24] B. A. Wujek, J. E. Renaud, S. M. Batill, and J. B. Brockman, "Concurrent subspace optimization using design variable sharing in a distributed computing environment," *Concurrent Engineering Research and Applications*, vol. 4, no. 4, pp. 361–376, 1996.

[25] J. Sobieszczanski-Sobieski, J. S. Agte, and R. R. J. Sandusky, "Bi-level integrated system synthesis (BLISS)," NASA Technical Memorandum 208715, NASA, Hampton, Va, USA, 1998.

[26] R. Braun, *Collaborative optimization: an architecture for large-scale distributed design [Ph.D. dissertation]*, Stanford University, Palo Alto, Calif, USA, 1996.

[27] N. Michelena, H. Park, and P. Y. Papalambros, "Convergence properties of analytical target cascading," *AIAA Journal*, vol. 41, no. 5, pp. 897–905, 2003.

[28] Y. Li, Z. Lu, and J. J. Michalek, "Diagonal quadratic approximation for parallelization of analytical target cascading," *Journal of Mechanical Design*, vol. 130, no. 5, pp. 051402-051401–051402-051411, 2003.

[29] S. Tosserams, L. F. P. Etman, P. Y. Papalambros, and J. E. Rooda, "An augmented lagrangian relaxation for analytical target cascading using the alternating directions method of multipliers," in *Proceedings of the 6th World Congress of Structural and Multidisciplinary Optimization*, Rio de Janeiro, Brazil, May 2005.

[30] G. Hardin, "The tragedy of the commons," *Science*, vol. 162, no. 3859, pp. 1243–1248, 1968.

[31] T. L. Vincent, "Game theory as a design tool," *Journal of Mechanisms, Transmissions, and Automation in Design*, vol. 105, no. 2, pp. 165–170, 1983.

[32] J. Eddy, *Solving distributed, non-cooperative design problems using multi agent systems [Ph.D. dissertation]*, University at Buffalo, Buffalo, NY, USA, 2006.

[33] K. Lewis and F. Mistree, "Collaborative, sequential, and isolated decisions in design," *Journal of Mechanical Design*, vol. 120, no. 4, pp. 643–652, 1998.

[34] C. Loch, J. Mihm, and A. Huchzermeier, "Concurrent engineering and design oscillations in complex engineering projects," *Concurrent Engineering Research and Applications*, vol. 11, no. 3, pp. 187–200, 2003.

[35] R. I. Whitfield, A. H. B. Duffy, G. Coates, and W. Hills, "Distributed design coordination," *Research in Engineering Design*, vol. 13, no. 3, pp. 243–252, 2002.

[36] S. Cho and S. D. Eppinger, "A simulation-based process model for managing complex design projects," *IEEE Transactions on Engineering Management*, vol. 52, no. 3, pp. 316–328, 2005.

[37] T. R. Browning and S. D. Eppinger, "Modeling impacts of process architecture on cost and schedule risk in product development," *IEEE Transactions on Engineering Management*, vol. 49, no. 4, pp. 428–442, 2002.

[38] E. Devendorf, M. Devendorf, and K. Lewis, "Using network theory to model distributed design systems," in *Proceedings of the 13th AIAA ISSMO Multidisciplinary Analysis and Optimization Conference (AIAA '10)*, 9027, Fort Worth, Tex, USA, 2010.

[39] K. Lewis and F. Mistree, "Collaborative, sequential and isolated decisions in design," in *Proceedings of the ASME Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Sacramento, Calif, USA, 1997.

[40] J. Nash, "Non-cooperative games," *Annals of Mathematics*, vol. 54, pp. 286–295, 1951.

[41] G. L. Thompson, "Signaling strategies in N-person games," in *Contributions to the Theory of Games*, H. W. Kuhn and A. W. Tucker, Eds., pp. 267–277, Princeton University Press, Princeton, NJ, USA, 1953.

[42] G. Hernandez, C. C. Seepersad, and F. Mistree, "Designing for maintenance: a game theoretic approach," *Engineering Optimization*, vol. 34, no. 6, pp. 561–577, 2002.

[43] K. Lewis and F. Mistree, "Modeling subsystem interactions: a game theoretic approach," *Journal of Design Manufacturing and Automation*, vol. 1, no. 1, pp. 17–26, 2001.

[44] R. P. Smith and S. D. Eppinger, "Identifying controlling features of engineering design," *Management Science*, vol. 43, no. 3, pp. 176–193, 1997.

[45] D. Poole, *Linear Algebra: A Modern Introduction*, Thomas Learning, 2003.

[46] K. Ogata, *Discrete-Time Control Systems*, Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 1995.

[47] V. Chanron, T. Singh, and K. Lewis, "Equilibrium stability in decentralized design systems," *International Journal of Systems Science*, vol. 36, no. 10, pp. 651–662, 2005.

[48] A. Gurnani and K. Lewis, "Collaborative, decentralized engineering design at the edge of rationality," *Journal of Mechanical Design*, vol. 130, no. 12, pp. 121101–121109, 2008.

[49] J. Kelly and M. Walker, "Critical-path planning and scheduling," in *Proceedings of the Eastern Joint Computer Conference*, Boston, Mass, USA, 1959.

[50] D. Malcolm, J. Roseboom, C. Clark, and W. Fazar, "Application of a technique for research and development program evaluation," *Journal of Operations Research*, vol. 7, no. 5, pp. 646–669, 1959.

[51] R. V. Slyke, "Monte carlo methods and the PERT problem," *Journal of Operations Research*, vol. 5, no. 11, pp. 839–860, 1963.

[52] A. Pritsker, *Modeling and Analysis Using Q-GERT Networks*, Wiley & Sons, New York, NY, USA, 2nd edition, 1979.

[53] S. Elmaghraby and J. Kamburowski, "The analysis of activity networks under generalized precedence relationships (GPR's)," *Journal of Management Science*, vol. 38, no. 9, pp. 1245–1263, 1992.

[54] J. Sobieszczanski-Sobieski, "On the sensitivity of complex, internally coupled systems," *AIAA Journal*, vol. 28, no. 1, pp. 173–180, 1990.

[55] J. L. Rogers, "DeMAID/GA an enhanced design manager's aid for intelligent decomposition," in *Proceedings of the 6th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization (AIAA '96)*, 4157, Seattle, Wash, USA, 1996.

[56] D. V. Steward, "The design structure system: a method for managing the design of complex systems," *IEEE Transactions on Engineering Management*, vol. 28, no. 3, pp. 71–74, 1981.

[57] S. D. Eppinger, D. E. Whitney, R. P. Smith, and D. A. Gebala, "A model-based method for organizing tasks in product development," *Research in Engineering Design*, vol. 6, no. 1, pp. 1–13, 1994.

[58] R. P. Smith and S. D. Eppinger, "A predictive model of sequential iteration in engineering design," *Management Science*, vol. 43, no. 8, pp. 1104–1120, 1997.

[59] A. Y. Ha and E. L. Porteus, "Optimal timing of reviews in concurrent design for manufacturability," *Management Science*, vol. 41, no. 9, pp. 1431–1447, 1995.

[60] V. Krishnan, S. D. Eppinger, and D. E. Whitney, "A model-based framework to overlap product development activities," *Management Science*, vol. 43, no. 4, pp. 437–451, 1997.

[61] B. A. Huberman and D. M. Wilkinson, "Performance variability and project dynamics," *Computational and Mathematical Organization Theory*, vol. 11, no. 4, pp. 307–332, 2005.

[62] J. Mihm, C. Loch, and A. Huchzermeier, "Problem-solving oscillations in complex engineering projects," *Management Science*, vol. 49, no. 6, pp. 733–750, 2003.

[63] A. Yassine, N. Joglekar, D. Braha, S. Eppinger, and D. Whitney, "Information hiding in product development: the design churn effect," *Research in Engineering Design*, vol. 14, no. 3, pp. 145–161, 2003.

[64] C. Sen, F. Ameri, and J. D. Summers, "An entropic method for sequencing discrete design decisions," *Journal of Mechanical Design*, vol. 132, no. 10, pp. 10100401–10100411, 2010.

[65] E. Devendorf and K. Lewis, "Are we there yet? Investigating the role of design process architecture in convergence time," in *Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (DETC '09)*, 87517, pp. 997–1008, San Diego, Calif, USA, September 2009.

[66] D. Braha and Y. Bar-Yam, "Topology of large-scale engineering problem-solving networks," *Physical Review E*, vol. 69, no. 1, Article ID 016113, pp. 161131–161137, 2004.

[67] D. Braha and Y. Bar-Yam, "The statistical mechanics of complex product development: empirical and analytical results," *Management Science*, vol. 53, no. 7, pp. 1127–1145, 2007.

[68] K. Ogata, *Modern Control Engineering*, Prentice Hall, New Delhi, India, 4th edition, 2005.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 763545, 10 pages
http://dx.doi.org/10.1155/2013/763545

Research Article

Research on Multidisciplinary Optimization Design of Bridge Crane

Tong Yifei,[1] Ye Wei,[1] Yang Zhen,[1] Li Dongbo,[1] and Li Xiangdong[2]

[1] *Nanjing University of Science and Technology, School of Mechanical Engineering 402, 210094 Nanjing, China*
[2] *Jiangsu Province Special Equipment Safety Supervision Inspection Institute, LongJiang Building, 107 Caochangmen Street, Jiangsu Province, 210000 Nanjing, China*

Correspondence should be addressed to Li Dongbo; db_tyf@yahoo.cn

Received 17 February 2013; Accepted 7 May 2013

Academic Editor: C. Wu

Bridge crane is one of the most widely used cranes in our country, which is indispensable equipment for material conveying in the modern production. In this paper, the framework of multidisciplinary optimization for bridge crane is proposed. The presented research on crane multidisciplinary design technology for energy saving includes three levels, respectively: metal structures level, transmission design level, and electrical system design level. The shape optimal mathematical model of the crane is established for shape optimization design of metal structure level as well as size optimal mathematical model and topology optimal mathematical model of crane for topology optimization design of metal structure level is established. Finally, system-level multidisciplinary energy-saving optimization design of bridge crane is further carried out with energy-saving transmission design results feedback to energy-saving optimization design of metal structure. The optimization results show that structural optimization design can reduce total mass of crane greatly by using the finite element analysis and multidisciplinary optimization technology premised on the design requirements of cranes such as stiffness and strength; thus, energy-saving design can be achieved.

1. Introduction

Empirical design is often used for the structure design of bridge crane, which determines the design parameters of bridge crane and furthermore improves the performance. The traditional design method cannot work out accurate performance data resulting in the safe coefficient of crane over the design requirements greatly, which leads to the waste of materials and energy consumption, and so forth [1].

At present, a simplified structure to reduce the weight and lightweight design-based heuristic algorithm is usually adopted to achieve energy saving, most of which focus on single structural design improvement. With the rapid development of finite element analysis (FEA) technique [2, 3], the traditional design method is gradually replaced by finite element analysis and design. There is quite a lot of finite element analysis software such as ANSYS, ABAQUS, and HyperWorks [4, 5]. However, purely from structural design, to reduce the weight of the crane has been very limited, and blindly to reduce the weight would be a security risk. On the other hand, crane is a complex system composed of many subsystems, among which there exist weak or strong coupling relationships. Thus, crane energy-saving design is a multidisciplinary coupling engineering problem involving structural design, mechanical transmission, and electrical control, which is not a simple superposition and permutations of various disciplines design. Therefore, it is of urgent need from multidisciplinary point of view of structure, mechanical transmission, and electrical control to study the system-level energy-saving design of crane.

The present work was carried out in order to obtain simulation data of the bridge crane. In the next section, the framework of multidisciplinary optimization design is proposed. In Section 3, FE model of double girder crane is developed using commercial program HyperWorks, and the loading and the results of finite element analysis are given and discussed. Topology optimization and size optimization are further carried out, and the results of metal structural optimization are analyzed. In Section 4, system-level multidisciplinary energy-saving optimization design of bridge crane is further carried out with transmission-level design results feedback to energy-saving optimization design of metal structure. Finally, research conclusions are summarized.

2. Multidisciplinary Optimization Design Framework for Bridge Crane Energy Saving

The presented research on crane multidisciplinary design technology for energy saving includes three levels, respectively: energy-saving design of metal structures, energy-saving transmission design, and energy-saving electrical system design. Energy-saving design of metal structure involves structural lightweight design and arch curve design of beam; energy-saving transmission system design involves dynamic loading, transmission efficiency, and components lightweight, and energy-saving electric system design involves power loss.

In addition, optimal design of lifting findings dynamic loading and components lightweight are feedback to structure lightweight design for further design optimization. Also, arch curve can reduce climbing energy consumption, thereby reducing motor power losses. The arch curve optimization results need feedback for electrical energy saving. The multidisciplinary optimization design can be illustrated as shown in Figure 1.

3. Energy-Saving Optimization Design of Metal Structure Level of Bridge Crane

3.1. Development of FE Model of Double Girder Crane. Take a bridge crane used in a practical project as the research object, which is a 50 t-31.5 m double girder crane whose material parameter and usage are as follows:

(i) material: ordinary carbon steel Q235;
(ii) length of the crane (l): 31.5 m;
(iii) maximum lifting height: 12 m;
(iv) hoisting speed: 7.8 m/min;
(v) moving speed of the car: 38.5 m/min;
(vi) moving speed of the cart: 87.3 m/min.

And according to the GBT 3811-2008 "crane design standard," the working level of car is M5, and the working level of cart is M6 [6].

3.2. Geometric Modeling of Double Girder Crane. According to the engineering drawing, geometric model of the bridge crane is established by PRO/E, whose structure components include the up and down plates of end girders, the side plates of end girders, up and down plates of main girders, the side plates of main girders, multiple belly boards, feet frame, and various connection boards. The simplified geometric model is shown as Figure 2.

3.3. Model Processing. Import the geometric model of bridge crane into HyperMesh and clear it. Owing to that each plate is thin, partition the plates with shell elements for finite element simulation analysis. The shell elements should be created on the middle surface of the geometry. A group of middle surfaces should be constructed by using "midsurface" panel. The imported model contains some connectivity error or some other defects, so the operations as follows should be carried out after importing file model.

(1) Delete the unsheared surfaces.
(2) Fill the gaps (repair the missed surfaces).
(3) Set the tolerance values of geometric cleaning.
(4) Combine the red free edges with "equivalence."
(5) Delete the repeated surfaces.

3.4. Mesh Partitioning of Double Girder Crane. Welds connections between each board are taken place of the rigid connections, and mesh elements are created on extraction midsurface [7]. The calculation capacity and calculation efficiency must be considered when mesh partitioning. The finer the elements meshed are, the more accordant the partitioned model is with the actual condition, while computing time and memory usage will be increased largely. After taking all the above factors into account synthetically, set the element size as 50 mm × 50 mm for finite element analysis. The spot welds are used to simulate the connections between the end and main girders [8].

Due to that the bridge crane structure is symmetrical, take half of the model as research object in order to reduce the computing time and memory usage. The FE model of bridge cranes is shown in Figure 3.

3.5. Loading and Static Analysis. Both end girders and main girders are processed as simply supported beams [9–11]. Loading is illustrated as in Figure 4.

Constraint loadings of the crane are described as follows:

the movement in x, y, z directions and the rotation in z direction of position 1 are restrained;

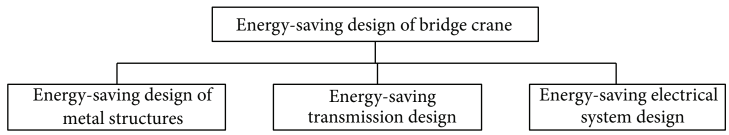
the movement in y, z direction and the rotation in z direction of position 4 are restrained;

the movement in z direction and the rotation in x, y direction of positions 2, 3 are restrained because of the symmetry.

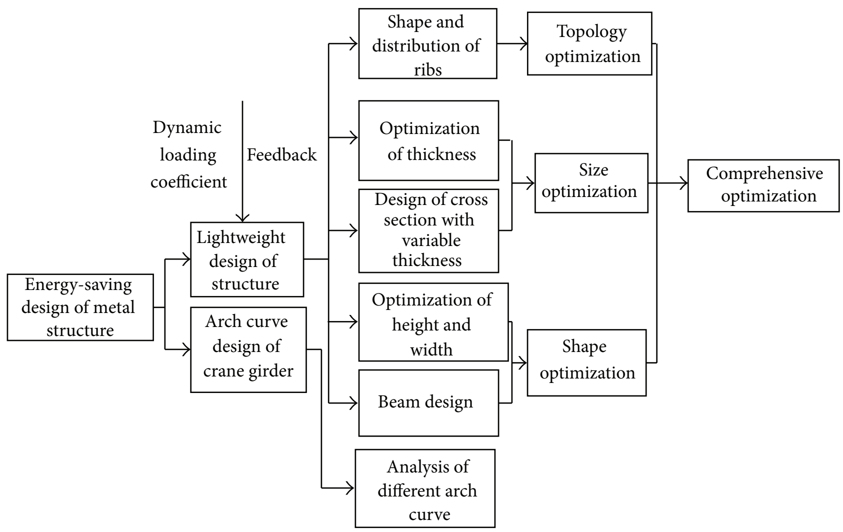
The loadings on both of the main girders are as follows:

(i) rated hoisting loading: $P_Q = 50$ t;
(ii) the car mass is 15.765 t;
(iii) self-vibration load factor $\Phi_1 = 1.1$;
(iv) lifting dynamic load factor $\Phi_2 = 1.14$;
(v) horizontal inertial force of crane as volume force: acceleration is 0.32 m/s^2;
(vi) cart gravity as volume force.

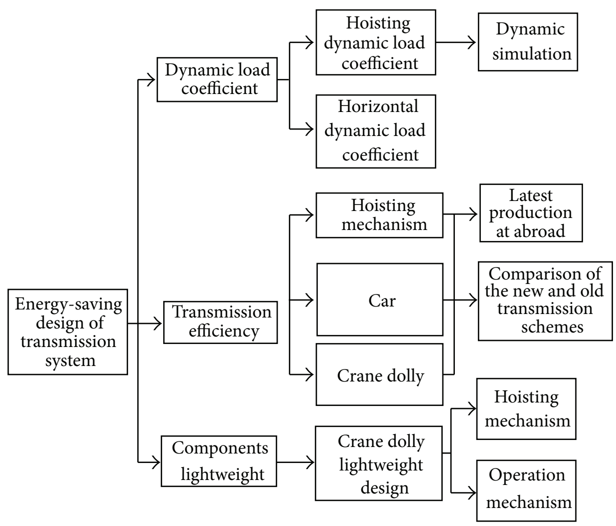
L0, L1, L2, L3, and L4 denote the loadings on different positions of one main girder, respectively, called five work



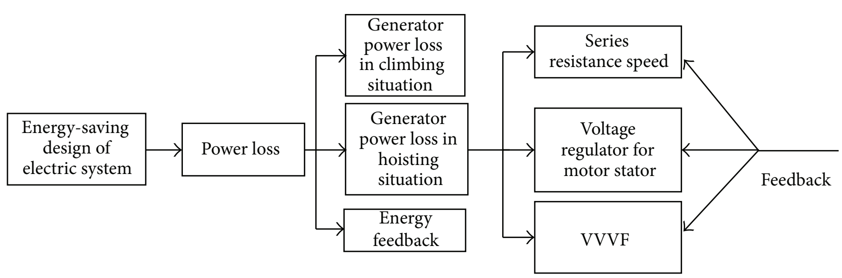
(a) Overall framework of multidisciplinary energy-saving design



(b) Energy-saving design of metal structure level



(c) Energy-saving design of transmission design level



(d) Energy-saving design of electrical system design level

Figure 1: Multidisciplinary optimization design framework for energy saving.

Table 1: Five work conditions description.

Work conditions	Cart gravity (m/s^2)	Car mass (t)	Self-vibration factor	Horizontal inertial force of crane (m/s^2)	Rated loading (t)	Lifting move load factor	Load position
1	9.8	15.765	1.1	0.32	50	1.14	Middle of the beam
2	9.8	15.765	1.1	0.32	50	1.14	Left end of the beam
3	9.8	15.765	1.1	0.32	50	1.14	Right end of the beam
4	9.8	15.765	1.1	0.32	50	1.14	Left 1/4 of the beam
5	9.8	15.765	1.1	0.32	50	1.14	Right 1/4 of the beam

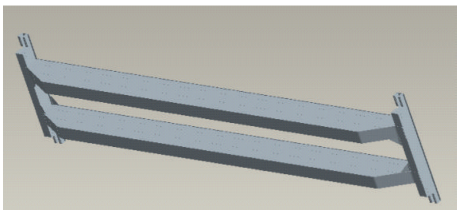


Figure 2: Geometric model of double girder crane.

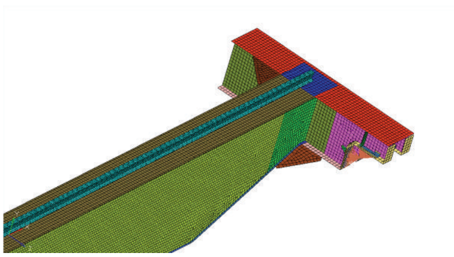


Figure 3: Finite element model.

conditions, and the magnitude of the loadings (L0, L1, L2, L3, and L4) is 322.2485 KN. Five work conditions are calculated in finite analysis as in Table 1.

According to the requirements of the crane design in GBT 3811-2008, "crane design" combined with actual usage, requirements for the stiffness of the crane girder are as follows:

$$f \leq \frac{1}{800} s, \quad (1)$$

where f is the deflection displacement, s is the span of the crane [12].

And requirements for the stress of the crane girder are as follows:

(i) material: Q235;

(ii) yield stress σ_s: 235 MPa;

(iii) allowable stress $[\sigma] \leq 100$ MPa defined by engineering design.

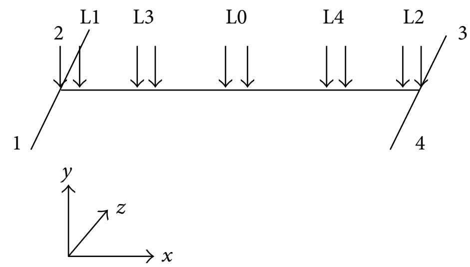


Figure 4: Loadings illustration.

After loading on different locations of the main girder, the results of finite element analysis are shown in Figures 5 and 6.

Analyz and compar different conditions of loads to obtain the conclusions that when loading on the middle of the main girder the maximum displacement of 40.3 mm appears on the middle of the main girder, and the maximum stress of 91.6 MPa occurs on the middle of the main girders. According to the results of FEA, the total mass of the initial model is 18.9 t.

3.6. *Structural Optimization of Double Girder Crane*

3.6.1. Shape Optimization. The shape optimal mathematical model is established as follows which takes the minimum volume as objective function, the height and width of the crane as design variables, and the scopes of stress, strain energy, and modal as constraints:

$$\begin{aligned} &\text{Min} \quad V'(X) = V'\left(\text{Height}', \text{Width}'\right) \\ &\text{Design variables:} \begin{matrix} -5 \leq \text{Height}' \leq 20 \\ -5 \leq \text{Width}' \leq 20 \end{matrix} \\ &C_j = \frac{1}{2} u_j^T f_j \leq 1.1 \times 10^7 J \quad j = 1, \ldots, 5 \\ &\text{s.t.} \quad Ku = f \\ &\qquad \sigma \leq 100\,\text{MPa} \\ &\qquad F \geq 3, \end{aligned} \quad (2)$$

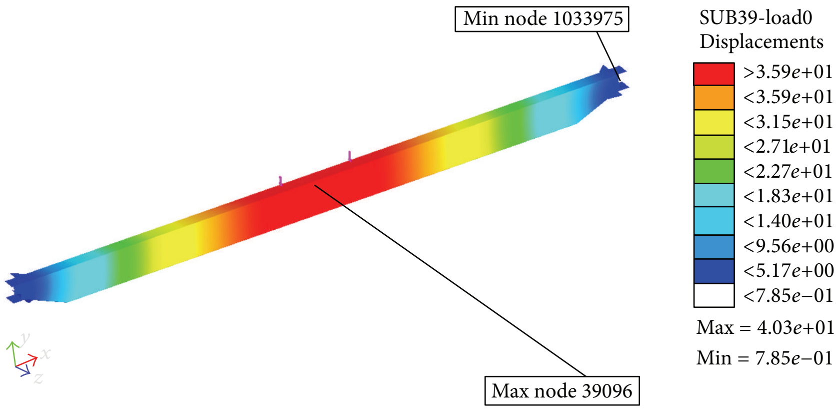


Figure 5: Displacement cloud.

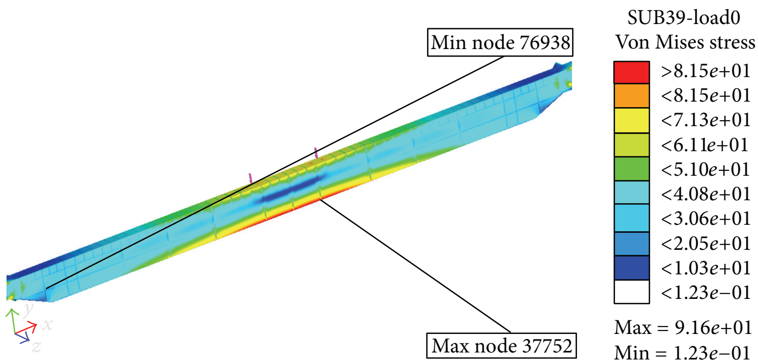


Figure 6: Stress cloud.

where $V'(X)$ denotes the volume fraction; C_j denotes the total strain energy of the crane under the jth load; K denotes the stiffness matrix of the system; f denotes the load; u denotes the node displacement vector under the load f; σ denotes the stress; F denotes the natural frequency. Objective function $V'(X)$, constraint function C_j, and σ can be obtained from structural response of the finite element analysis.

Use OptiStruct Solver to optimize the girder by selecting morph optimization tool; the optimization results of the main girder are shown as follows:

$$\text{Volume} = 2.27E + 09\,\text{mm}^3,\ \text{Mass} = 17.9\,\text{t}$$
$$\text{Height}' = -0.36,\ \text{Width}' = 1.2.$$

After shape optimization:

$$\text{Height} = 1724\,\text{mm} + 0.36 \times 50\,\text{mm} = 1742\,\text{mm}$$
$$\text{Width} = 600\ \text{mm} - 1.2 \times 50\,\text{mm} = 540\,\text{mm}$$
$$\text{Height/Width} = 1724/540 = 3.19.$$

By analyzing the results of finite element analysis, the structure performance (including strength, stiffness, and modal) after topology optimization meets the requirements of crane design specifications greatly, which are shown in Figures 7 and 8.

Analyze and compare different conditions of loads to obtain the conclusions that when loading on the middle of the main girder, the maximum displacement of 42.9 mm appears on the middle of the main girder and the maximum stress of 98.6 MPa occurs on the middle of the main girders. The total mass of the model after shape optimization is 17.9 t, which has reduced by 5.3%.

Compare the maximum displacement and maximum stress before and after topology optimization the result is given as in Table 2.

The analysis results shown in Table 2 show that structure performance of the various plates, some materials of which have been reasonably removed, meets the design requirements as well. Meanwhile, the total mass of structure is 17.9 t, which has reduced by 1 t.

3.6.2. Size Optimization. Furth optimizing of the structure after shape optimization was carried out in our research. Taking the minimum volume as the objective function, the thicknesses of the plates as the design variables, the scopes of the stress, strain energy, and modal as constraints, the size optimal mathematical model is established as follows:

$$\text{Min} \quad V(X) = V(x_1, x_2, \ldots, x_{18})$$
$$\text{S.T.} \quad C_j = \frac{1}{2} u_j^T f_j \le 1.1 \times 10^7 J \quad j = 1, \ldots, 5$$

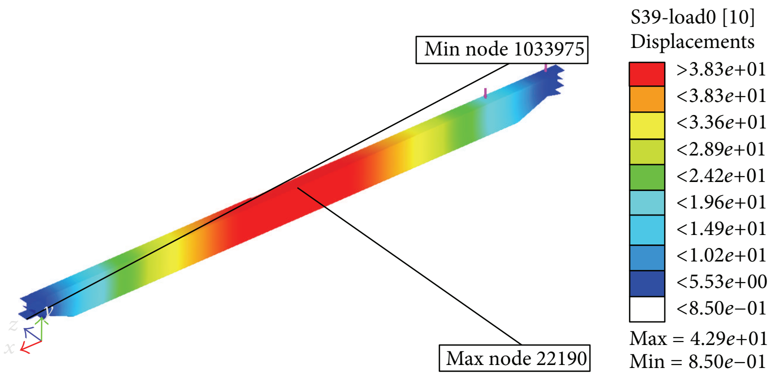


FIGURE 7: Displacement cloud.

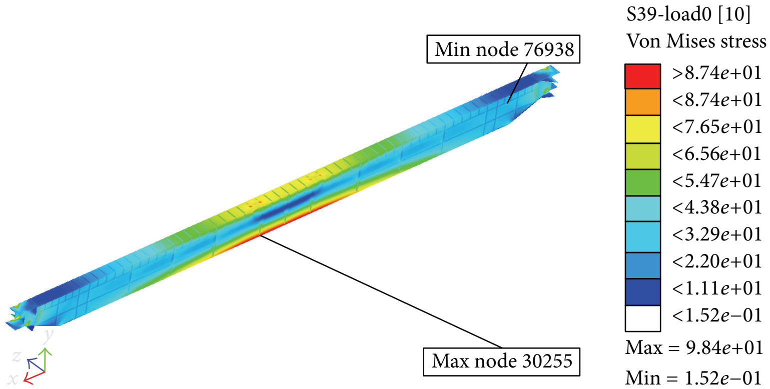


FIGURE 8: Stress cloud.

$$\begin{aligned} &Ku = f \\ &\sigma \leq 100\,\text{MPa} \\ &F \geq 3, \end{aligned} \tag{3}$$

where $X = x_1, x_2, \ldots, x_{18}$ denotes the thicknesses of plates, $V(X)$ denotes the total volume of the crane, the rest of variable parameters are denoted as above. Use the OptiStruct Solver to optimize girders by size optimization tool. The optimization results of the thicknesses of the plates are shown in Table 3.

By analyzing the results of size optimization, the structure performance (including strength, stiffness, and modal) after topology optimization meets the requirements of crane design specifications greatly. The results of finite element analysis after size optimization are shown in Figures 9 and 10.

The maximum displacement of 44.1 mm appears on the middle of the main girder, and the maximum stress of 99 MPa occurs on the end of the main girders. The total mass of the model after size optimization is 17.3 t, which has reduced by 8.5%.

The comparison of initial model and final model is shown in Table 4.

From the analysis results shown in Table 4, it can be found easily that the structure performance after shape and size optimization meets the requirements of crane design specifications greatly. Moreover, after size optimization, the total mass of the main girder changes into 17.3 t which has been reduced by 1.6 t.

3.6.3. Topology Optimization. Furth optimizing of the structure after shape and size optimization was carried out. The topology optimal mathematical model is established as follows which takes the minimum volume fraction as objective function, the material density of each element as design variables, and the scopes of stress, strain energy, and modal as constraints:

$$\begin{aligned} \text{Min} \quad & V'(X) = V'(x_1, x_2, \ldots, x_n) \\ & C_j = \frac{1}{2} u_j^T f_j \leq 1.1 \times 10^7 J \quad j = 1, \ldots, 5 \\ \text{s.t.} \quad & Ku = f \\ & \sigma \leq 100\,\text{MPa} \\ & F \geq 3, \end{aligned} \tag{4}$$

where $X = x_1, x_2, \ldots, x_n$ denotes the material density of each element are and the rest of variable parameters are denoted as above.

Table 2: Comparison of the stress and displacement.

Load step	Before		After	
	Stress (MPa)	Displacement (mm)	Stress (MPa)	Displacement (mm)
L0	91.6	40.3	98.6	42.9
L1/L2	80.2		87.2	
L3/L4	74.5		74.4	

Table 3: Comparison of the thicknesses before and after optimization.

Main optimal size	Before (mm)	After (mm)
Upper plates	24	20.7
Under plates	24	23.4
Inside plates	6	6
Outside plates	6	6
Small ribbed plates	8	9.5
Big ribbed plates	8	5.6

Use OptiStruct Solver to optimize the girder by selecting topology optimization tool; the optimization results of the main girder are shown in Figures 11 and 12.

By analyzing the results of topology optimization, the structure performance (including strength, stiffness and modal) after topology optimization meets the requirements of crane design specifications greatly. The results of finite element analysis after topology optimization are shown in Figures 13 and 14.

Analyze and compare different conditions of loads to obtain the conclusions that when loading on the middle of the main girder, the maximum displacement of 39.9 mm appears on the middle of the main girder and the maximum stress of 99 MPa occurs on the end of the main girders. The total mass of the model after topology optimization is 15.8 t, which has reduced by 16.4%.

Compare the maximum displacement and maximum stress before and after topology optimization; the result is given as in Table 5.

The analysis results shown in Table 5 show that structure performance of the various plates, some materials of which have been reasonably removed, meets the design requirements as well. Meanwhile, the total mass of structure is 15.8 t, which has reduced by 3.1 t.

3.7. Overall Stability Analysis of Main Girder. According to the requirements of "the crane design manual" for box section structure, "when aspect ratio (height/width) denoted by $h/b \leqslant 3$ or $3 < h/b \leqslant 6$ & $l/b \leq 95(235/\sigma_s)$, the lateral buckling stability of the flexural components do not need verify."

In our research, the results are as follows:

before optimization: $h = 1724$ mm, $b = 600$ mm, $h/b = 2.87$, so $h/b \leqslant 3$;

after optimization: $h = 1742$ mm, $b = 540$ mm, $h/b = 3.19$, and $l = 31500$ mm, $l/b = 58.3$, $\sigma_s = 253$ MPa, so $3 < h/b \leqslant 6$ & $l/b \leq 95(235/\sigma_s)$.

Therefore, lateral buckling stability conforms to the design requirements.

4. System-Level Multidisciplinary Energy-Saving Optimization Design of Bridge Crane

Energy-saving transmission design is researched by our research group in dynamic simulation and speed regulation of hoisting mechanism as well as optimization and innovation of transmission mechanism scheme reported in the literature [13, 14]. Thus, self-vibration load factor in Section 3.4 is reduced from 1.14 to 1.11 under VVVF, and the car mass in Section 3.4 is reduced from 15.765 t to 14.4 t. System-level multidisciplinary energy-saving optimization design of bridge crane can be further carried out with energy-saving transmission design results feedback to energy-saving optimization design of metal structure. By repeating the above modelling and analysis in Section 3, the system-level multidisciplinary energy-saving optimization results are shown in Table 6.

5. Conclusions

The framework of multidisciplinary energy-saving optimization design of bridge crane is proposed. And the structure-level optimization design of bridge crane by using finite element analysis technology is discussed in this paper in detail. This research seeks to get more reasonable, lightweight, and energy-saving structure on the basis of insuring the performances of crane and to provide the design reference for bridge crane. The main results of this research can be concluded as follows:

(1) the results of finite element analysis show that the concentrated stress occurs on the middle of main girders under full load;

(2) for the cranes which meet the design requirements, shape optimization is researched. The total mass of the structure after shape optimization changes into 17.9 t/17.6 t (optimization design of metal structure/system-level multidisciplinary energy-saving optimization), and it is reduced by 1 t/1.3 t compared with the initial model;

(3) size optimization is researched after shape optimization. The total mass of the structure after size optimization changes into 17.3 t/16.7 t, and it is reduced by 1.6 t/2.2 t;

Table 4: Comparing the stress and displacement.

Load step	Initial model		Final model	
	Stress (MPa)	Displacement (mm)	Stress (MPa)	Displacement (mm)
L0	91.6	40.3	99	44.1
L1/L2	80.2		84.4	
L3/L4	74.5		88.3	

Table 5: Comparison of the stress and displacement.

Load step	Before		After	
	Stress (MPa)	Displacement (mm)	Stress (MPa)	Displacement (mm)
L0	91.6	40.3	97.6	43.9
L1/L2	80.2		90	
L3/L4	74.5		108	

Table 6: System-level multidisciplinary energy-saving optimization results.

Optimization method	Mass after optimization (t)	Percentage decrease
	18.9	
Shape optimization	17.6	6.88%
Size optimization	16.7	11.64%
Topology optimization	15.8	16.40%

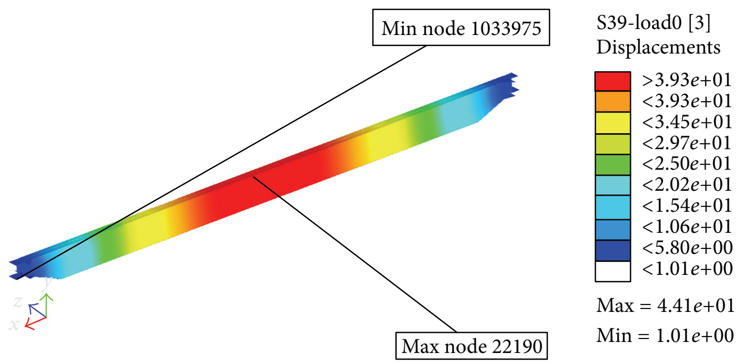


Figure 9: Displacement cloud.

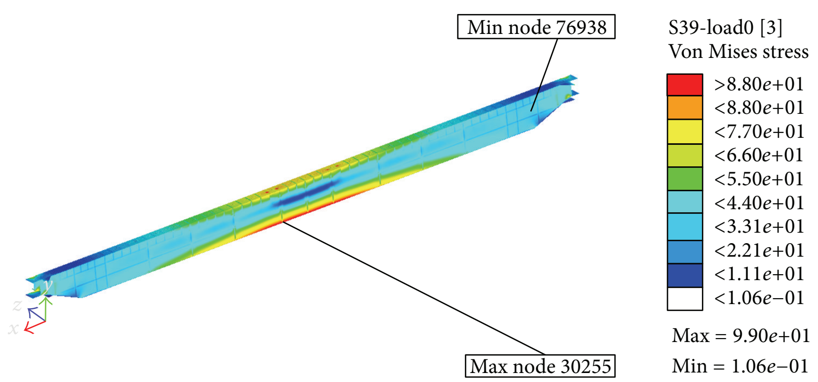


Figure 10: Stress cloud.

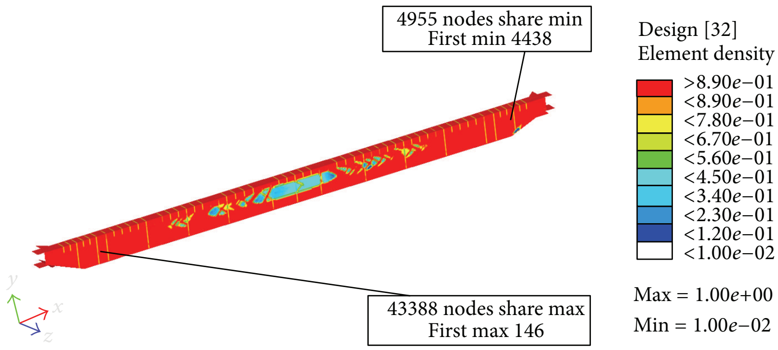


FIGURE 11: Density graph of the side plate.

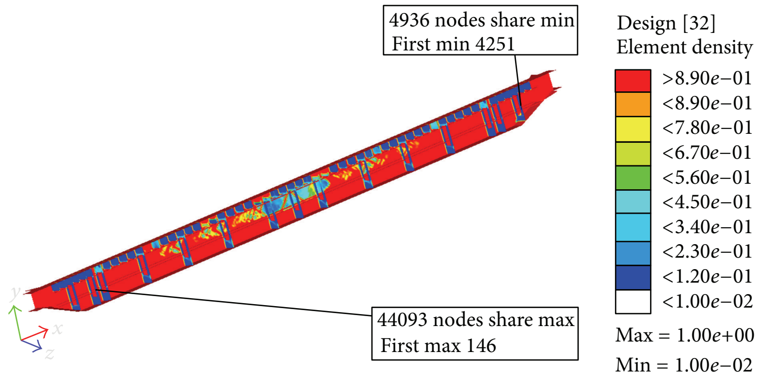


FIGURE 12: Density graph of the belly board.

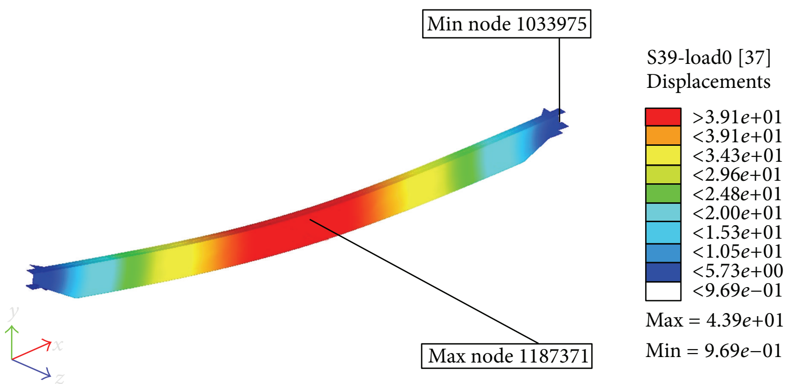


FIGURE 13: Displacement cloud.

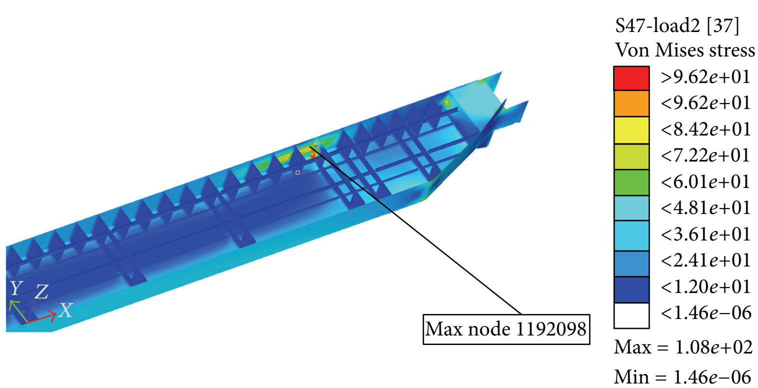


FIGURE 14: Stress cloud.

(4) topology optimization based on density methodology is used after shape and size optimization. The total mass of the structure after topology optimization changes into 15.8 t/15.8 t, and it is reduced by 3.1 t/3.1 t compared with the initial model;

(5) multidisciplinary optimization design by means of finite element analysis and dynamic simulation not only can assure stiffness, strength, and other performances requirements of the crane but also can greatly reduce the use of materials by lightweight design.

Acknowledgments

This work was financially supported by National Foundation of General Administration of Quality Supervision and Inspection (2012QK178), program of Science Foundation of General Administration of Quality Supervision and Inspection of Jiangsu Province (KJ103708), and "excellence plans-zijin star" Foundation of Nanjing University of Science. The support is gratefully acknowledged.

References

[1] N. Zhaoyang, R. Yongxin, R. Chengao et al., "ANSYS application in design of the main beam of crane," *Machinery for Lifting and Transportation*, vol. 5, pp. 31–33, 2008.

[2] M. Styles, P. Compston, and S. Kalyanasundaram, "Finite element modelling of core thickness effects in aluminium foam/composite sandwich structures under flexural loading," *Composite Structures*, vol. 86, no. 1–3, pp. 227–232, 2008.

[3] C. Nucera and F. L. Scalea, "Higher-harmonic generation analysis in complex waveguides via a nonlinear semianalytical finite element algorithm," *Mathematical Problems in Engineering*, vol. 2012, Article ID 365630, 16 pages, 2012.

[4] L. Kwasniewski, H. Li, J. Wekezer, and J. Malachowski, "Finite element analysis of vehicle-bridge interaction," *Finite Elements in Analysis and Design*, vol. 42, no. 11, pp. 950–959, 2006.

[5] Y. Li, Y. Pan, J. Zheng et al., "Finite element analysis with iterated multiscale analysis for mechanical parameters of composite materials with multiscale random grains," *Mathematical Problems in Engineering*, vol. 2011, Article ID 585624, 19 pages, 2011.

[6] D. Qin, Y. Wang, X. Zhu, J. Chen, and Z. Liu, "Optimized design of the main beam of crane based on MSC1Patran/Nastran," *Machinery for Lifting and Transportation*, vol. 7, pp. 24–26, 2007.

[7] D. Hongguang, B. Tianxiang, S. Yanzhong, and Y. Sikun, "The optimum design of single girder bridge-crane based on FEM," *Steel Construction*, vol. 2, pp. 46–48, 2009.

[8] L. Hui, Y. Haipeng, and L. Huixin, "The malfunction about the bridge crane girder based on FEM," *Construction Machinery*, vol. 2, pp. 67–69, 2007.

[9] W. Fumian, "Discussion about the static stiffness design of the bridge crane and the related problems," *Machinery for Lifting and Transportation*, vol. 12, pp. 42–44, 2009.

[10] Z. Zhang, "The application of ANSYS into optimizing design of main beam in joist portal crane," *Machinery Technology and Management on Construction*, vol. 8, pp. 91–93, 2009.

[11] Y. Menglin, Q. Dongchen, L. Zhuli, and W. Y. Jia, "Research on structure optimization design of bridge crane box beam," *Design and Research*, vol. 4, pp. 23–24, 2008.

[12] GBT3811-2008, Crane design standard.

[13] L. Shuishui, F. Yuanxun, and B. Tingchun, "Introduction for new hoisting mechanism of a crane," *Machinery Design & Manufacture*, no. 5, pp. 275–276, 2012.

[14] F. Yuanxun, B. Tingchun, and L. Shuishui, "Co-simulating on lifting dynamic load of bridge crane based on ADAMS and MATLAB," *Heavy Machinery*, no. 5, pp. 30–32, 2011.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 803609, 11 pages
http://dx.doi.org/10.1155/2013/803609

Research Article

An Analytical Method for Evaluating the Dynamic Response of Plates Subjected to Underwater Shock Employing Mindlin Plate Theory and Laplace Transforms

Zhenyu Wang, Xu Liang, and Guohua Liu

College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China

Correspondence should be addressed to Xu Liang; liangxu@zju.edu.cn

Received 2 March 2013; Accepted 25 April 2013

Academic Editor: Zhuming Bi

It is often in the interest of a designer to know the transient state of stress in a plate subjected to an underwater explosion. In this paper, an analytical method based on Taylor's fluid-solid interaction (FSI) model, Mindlin plate theory, Laplace transform, and its inversion is proposed to examine the elastic dynamic response of a plate subjected to an underwater explosion. This analytical method includes shear deformation, the moments and membrane stress in the plate, and the FSI effect and considers a full profile of possibilities. The results of the response-time histories and the response distribution on the plate in terms of displacements and stresses from the analytical method are compared with finite element analysis (FEA) to validate this method, and the comparison indicates good agreement. Comparison of the acceleration at the center of an air-backed plate between the analytical method and the experiment from relevant literature, shows good agreements, and the analytical method and its FSI model are validated. The influence of the FSI is investigated in detail. All extreme values of the response-time histories decrease as the thickness increases for the non-FSI case. The results can be used as benchmark solutions in further research.

1. Introduction

Plate element is one of the basic elements of every class of ship and apparatus and may be subjected to underwater explosions. It is in the interest of a ship designer to know the transient state of stress in advance that develops in a plate during an underwater explosion. For low-intensity explosions, the stresses developed within a plate are usually within the material's elastic range [1–3].

The classical plate theory (CPT) has been used to solve the response in thin plates to an underwater explosion. An orthotropic and simply supported plate reinforced by stiff particles embedded in a matrix offers the potential for simple, economically functional grading and enhanced response. Genin and Birman [4] obtained a solution for the response of this plate under a uniform and time-varying overpressure using the Kirchhoff plate theory and a convolution integral. However, the shear deformation through the thickness is neglected within CPT, and the effect of fluid-solid interaction (FSI) was not considered in this research.

Some of recent research on this subject have been devoted to the extension of linear to nonlinear response by involving von Kármán thin plate theory and employing the extended Galerkin and the Runge-Kutta methods. Hause [5] investigated the nonlinear response of a functionally graded plate with two constituent phases exposed to a Friedlander explosive air blast within the classical plate theory, but the effect of fluid-solid interaction (FSI) was not considered in this research. Librescu et al. [6–9] have developed a three-dimensional sandwich model which is used in the investigation of the dynamic responses of a sandwich plate subjected to underwater explosions by employing Hayman's FSI model [6, 7]. In the sandwich model, core layer can only carry transverse shear stresses, while the transverse shear effects are neglected for the face sheets. In general, all of above research concerning the nonlinear response are numerical but not analytical.

The phenomenon of FSI significantly influences the dynamic response of a structure during an underwater explosion. The influence of FSI is typically investigated by finite element analysis (FEA) which provides a detailed picture depending on the level of complexity of the model [10–13]. The discretization used both for spatial derivation and time integration, however, affects spurious oscillations and the accuracy of a numerical solution [14].

Taylor [15] proposed a one-dimensional FSI model for a planar shock wave impinging on a freestanding plate and considered the momentum transmitted to the plate by the shock wave. Liu and Young [16] extended Taylor's air-backed FSI model to a water-backed FSI model and systematically investigated the influence of the back conditions on the interaction between the fluid and the structure. Utilizing Taylor's FSI model, the energy method, and considering the conditions of the air-backed and water-backed cases, Rajendran et al. [1–3] derived one-dimensional, semianalytical models to predict the elastic strains in circular and rectangular plates subjected to underwater explosions. Good agreement was found between the model and the experimental results. However, these solutions considering the FSI effect are one-dimensional and ignore moment and membrane stress in the plate. In addition to elastic method, plastic dynamic analytical solutions for the response of a three-layered sandwich plate with a soft core to UNDEX have also formed the subject of recent studies [11, 17–20], in which the existing plastic solution procedure generally comprises three stages. However, the velocity is assumed to be uniform for the front and back faces of the sandwich plate except for the boundaries in the FSI phase, with the FSI differences between the different positions on front and back faces of plate being ignored, that is to say, only one-dimensional FSI effect is considered.

A comprehensive review on the methods for the response of plates to underwater explosion is present above. It is worthy noted that the analytical methods mentioned above do not consider the FSI differences between the different positions on front and back faces of plate. The purpose of this paper is to fill this gap by proposing an analytical method to determine the elastic dynamic response of a plate subjected to an underwater explosion based on Taylor's FSI model, Mindlin plate theory, Laplace transform, and its inversion. This method considers the full profile of possibilities, such as non-FSI, air-backed, and water-backed cases. Section 1 introduces the solution methods. In Section 2, the problem is described and Taylor's FSI model, the air-backed, and the water-backed cases are introduced. Section 3 introduces the governing equations of motion for Mindlin plate theory. In Section 4, analytical solutions for the three cases are derived by Fourier transform, Laplace transform, and their inversion. In Section 5, the analytical results are validated by being compared with those from the FEA and the experiment record, and the influences of the FSI and material thickness are investigated. The conclusions are discussed in Section 6.

2. Problem Description and Fluid-Solid Interaction

Consider a rectangular plate of length a, width b, and a uniform thickness h, as shown in Figure 1. The plate is

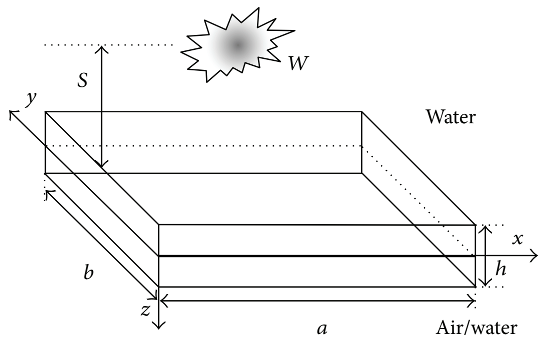


Figure 1: A rectangular plate subjected to an underwater explosion showing dimensions and the Cartesian coordinate system.

subjected to an underwater explosion from an explosive charge W (TNT equivalent weight in kg) located at a distance S (in m) from the center of the plate.

Taylor [15] proposed a one-dimensional FSI model for a planar shock wave impinging on a free-standing plate, including the momentum transmitted to the plate by the shock wave. This FSI model has been widely used in analytical methods modeling the plastic response of plates subjected to an underwater explosion [17, 19, 20] and is briefly reviewed below.

Adopting the planar wave assumption, the water pressure $p_f(x, y, t)$ at the front face ($z = -h/2$) comprises the incident wave $p_i(x, y, t)$, the idealized perfectly reflected wave $p_{r1}(x, y, t)$, and the rarefaction wave pressure $p_{r2}(x, y, t)$, respectively, as follows [15]:

$$\begin{aligned} p_f(x, y, t) &= p_i(x, y, t) + p_{r1}(x, y, t) + p_{r2}(x, y, t) \\ &= 2p_m \exp\left(-\frac{t}{t_d}\right) - \rho_w c_w \dot{w}\left(x, y, z = -\frac{h}{2}, t\right), \end{aligned} \tag{1}$$

where w denotes the deflection of the plate and the dots over a quantity stand for time derivatives.

The peak pressure p_m and the time parameter t_d of the incident wave are given as follows [1, 21]:

$$\begin{aligned} p_m &= 52.16 \times \left(\frac{W^{1/3}}{S}\right)^{1.13} \text{MPa}, \\ t_d &= 96.5 \times 10^{-3} \times W^{1/3} \times \left(\frac{W^{1/3}}{S}\right)^{-0.22} \text{msec}. \end{aligned} \tag{2}$$

Two backing condition categories are water-backed and air-backed. For an air-backed plate, the water pressure at the back face ($z = h/2$) is [1]

$$p_b(x, y, t) = 0, \tag{3}$$

and for a water-backed plate, the pressure at the back face is [1]

$$p_b(x, y, t) = -\rho_w c_w \dot{w}\left(x, y, z = \frac{h}{2}, t\right). \tag{4}$$

To compare the two cases with the FSI effects between the plate and the water, a non-FSI case (ignoring FSI effects) is also discussed. The water pressures at both front and back faces can be expressed as follows [1]:

$$p_f(x,y,t) = p_m \exp\left(-\frac{t}{t_d}\right), \quad p_b(x,y,t) = 0. \tag{5}$$

3. Fundamental Equations

Considering an orthotropic material, the Mindlin plate theory implies that the displacements are distributed in the plate as follows:

$$u = -z\psi_{xz}, \quad v = -z\psi_{yz}, \quad w = w_0, \tag{6}$$

where w_0 is the vertical displacement at the middle surface.

The moment and shear force resultants can be written in terms of the displacements as

$$\begin{bmatrix} M_x \\ M_y \\ M_{xy} \end{bmatrix} = \frac{h^3}{12}\begin{bmatrix} c_{11} & c_{12} & 0 \\ c_{12} & c_{22} & 0 \\ 0 & 0 & c_{66} \end{bmatrix}\begin{bmatrix} -\psi_{xz,x} \\ -\psi_{yz,y} \\ \dfrac{-\left(\psi_{xz,y}+\psi_{yz,x}\right)}{2} \end{bmatrix}, \quad \begin{bmatrix} Q_y \\ Q_x \end{bmatrix} = \frac{\kappa h}{2}\begin{bmatrix} c_{44} & 0 \\ 0 & c_{55} \end{bmatrix}\begin{bmatrix} w_{0,y}-\psi_{yz} \\ w_{0,x}-\psi_{xz} \end{bmatrix}, \tag{7}$$

where ψ_{xz} and ψ_{yz} are the rotations about the x- and y-axes, respectively, h is the thickness of the plate, κ is the shear correction factor (taken as 5/6), and $c_{11}, c_{12}, c_{22}, c_{44}, c_{55}$, and c_{66} are the elastic constants. Throughout the paper, a comma after a quantity denotes a particular differentiation with respect to the spatial coordinates.

The normal stresses are distributed as

$$\sigma_{xx} = \frac{12zM_x}{h^3}, \quad \sigma_{yy} = \frac{12zM_y}{h^3}, \quad \sigma_{zz} = 0, \tag{8}$$

and the shear stresses are distributed as

$$\sigma_{xy} = \frac{12zM_{xy}}{h^3}, \quad \sigma_{xz} = \frac{3\left(1-4z^2/h^2\right)Q_x}{(2h)}, \quad \sigma_{yz} = \frac{3\left(1-4z^2/h^2\right)Q_y}{(2h)}. \tag{9}$$

The von Mises stress σ_m can be expressed as

$$\sigma_m = \sqrt{\frac{\left[\left(\sigma_{xx}-\sigma_{yy}\right)^2 + \left(\sigma_{xx}-\sigma_{zz}\right)^2 + \left(\sigma_{yy}-\sigma_{zz}\right)^2 + 6\left(\sigma_{xy}^2+\sigma_{xz}^2+\sigma_{yz}^2\right)\right]}{2}}. \tag{10}$$

The governing equations of motion for the Mindlin plate theory are

$$M_{x,x} + M_{xy,y} + \rho_p J\ddot{\psi}_{xz} = Q_x, \quad M_{xy,x} + M_{y,y} + \rho_p J\ddot{\psi}_{yz} = Q_y, \quad Q_{x,x} + Q_{y,y} - \rho_p h\ddot{w}_0 + p_f + p_b. = 0, \tag{11}$$

where $J = h^3/12$.

Substituting (7) into (11) allows (11) to be rewritten as

$$\frac{h^3\left[2\left(c_{11}\psi_{xz,xx}+c_{12}\psi_{yz,xy}\right)+c_{66}\left(\psi_{xz,yy}+\psi_{yz,xy}\right)\right]}{24} - \frac{h^3\rho_p\ddot{\psi}_{xz}}{12} + \frac{c_{55}\kappa h\left(w_{0,x}-\psi_{xz}\right)}{2} = 0, \tag{12}$$

$$\frac{h^3\left[2\left(c_{12}\psi_{xz,xy}+c_{22}\psi_{yz,yy}\right)+c_{66}\left(\psi_{xz,xy}+\psi_{yz,xx}\right)\right]}{24} - \frac{h^3\rho_p\ddot{\psi}_{yz}}{12} + \frac{c_{44}\kappa h\left(w_{0,y}-\psi_{yz}\right)}{2} = 0, \tag{13}$$

$$\frac{\rho_p h\ddot{w}_0 - \kappa h\left[c_{55}\left(w_{0,xx}-\psi_{xz,x}\right)+c_{44}\left(w_{0,yy}-\psi_{yz,y}\right)\right]}{2} = p_f + p_b. \tag{14}$$

Assuming the plate is simply supported, the edge conditions are given as

$$x = 0, \quad a: w_0 = \psi_{yz} = M_x = 0, \quad y = 0, \quad b: w_0 = \psi_{xz} = M_y = 0. \tag{15}$$

The following Fourier series expansions for the displacements satisfy (15):

$$w_0 = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\overline{w}_0 \sin\alpha_m x \sin\beta_n y, \quad \psi_{xz} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\overline{\psi}_{xz}\cos\alpha_m x\sin\beta_n y, \quad \psi_{yz} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\overline{\psi}_{yz}\sin\alpha_m x\cos\beta_n y, \tag{16}$$

where

$$\alpha_m = \frac{m\pi}{a}, \qquad \beta_n = \frac{n\pi}{a}, \tag{17}$$

and $m, n = 1, 2 \ldots$ are the number of half waves in the x and y directions, respectively, and the overscript $(\bar{\ })$ denotes the transformed function in the Fourier domain.

According to the plane wave assumption, the incident wave and the idealized perfectly reflected wave can also be expanded in Fourier series form as

$$p_i = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \overline{p}_i \sin\alpha_m x \sin\beta_n y, \tag{18}$$

$$\overline{p}_i = \frac{4}{ab}\int_0^a\int_0^b p_i \sin\alpha_m x \sin\beta_n y dx\, dy = \frac{4\eta_1 p_i}{ab}, \tag{19}$$

$$p_{r1} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \overline{p}_{r1} \sin\alpha_m x \sin\beta_n y, \tag{20}$$

$$\overline{p}_{r1} = \frac{4}{ab}\int_0^a\int_0^b p_{r1} \sin\alpha_m x \sin\beta_n y dx\, dy = \frac{4\eta_1 p_{r1}}{ab}, \tag{21}$$

where

$$\eta_1 = \frac{\left[-1+(-1)^m\right]\left[-1+(-1)^n\right]}{(mn\pi^2)}. \tag{22}$$

4. Solutions Procedure

The analytical solutions in the Fourier and Laplace transformed domains are derived for the three cases, respectively. The final solutions to the displacement distributions are obtained by the inversion of the Fourier and Laplace transforms. The inversion of Laplace transform is carried out analytically.

4.1. Non-FSI. For the non-FSI case, by substituting (5) and (16)–(19) into (12)–(14), the fundamental equation can be rewritten in the Fourier transformed domain as follows:

$$\begin{aligned}&\frac{h^3\rho_p\ddot{\overline{\psi}}_{xz}}{12} + \overline{\psi}_{xz}\left[\frac{h^3\left(2c_{11}\alpha_m^2 + c_{66}\beta_n^2\right)}{24} + \frac{c_{55}h\kappa}{2}\right]\\&\quad + \frac{(2c_{12}+c_{66})h^3\alpha_m\beta_n\overline{\psi}_{yz}}{24} - \frac{c_{55}h\kappa\overline{w}_0\alpha_m}{2} = 0,\end{aligned} \tag{23}$$

$$\begin{aligned}&\frac{h^3\rho_p\ddot{\overline{\psi}}_{yz}}{12} + \overline{\psi}_{yz}\left[\frac{h^3\left(c_{66}\alpha_m^2 + 2c_{22}\beta_n^2\right)}{24} + \frac{c_{44}h\kappa}{2}\right]\\&\quad + \frac{(2c_{12}+c_{66})h^3\alpha_m\beta_n\overline{\psi}_{xz}}{24} - \frac{c_{44}h\kappa\overline{w}_0\beta_n}{2} = 0,\end{aligned} \tag{24}$$

$$\begin{aligned}&h\rho_p\ddot{\overline{w}}_0 + \frac{h\kappa\overline{w}_0\left(c_{55}\alpha_m^2 + c_{44}\beta_n^2\right)}{2} - \frac{c_{55}h\kappa\alpha_m\overline{\psi}_{xz}}{2}\\&\quad - \frac{c_{44}h\kappa\beta_n\overline{\psi}_{yz}}{2} = \frac{4\eta_1 p_m \exp(-t/t_d)}{ab}.\end{aligned} \tag{25}$$

The Laplace transform of the function $f(t)$ is defined as

$$L\left[f(t)\right] = \tilde{f}(s) = \int_0^{\infty} f(t)e^{-st}dt, \tag{26}$$

where s is the parameter of the Laplace transform and the overscript ($\tilde{\ }$) denotes the transformed function in the Laplace domain. Applying the Laplace transform to (23)–(25) yields:

$$\begin{aligned}&\left[\frac{h^3\left(2c_{11}\alpha_m^2 + c_{66}\beta_n^2 + 2s^2\rho_p\right)}{24} + \frac{c_{55}h\kappa}{2}\right]\tilde{\overline{\psi}}_{xz}\\&\quad + \frac{h^3\alpha_m\beta_n(2c_{12}+c_{66})\tilde{\overline{\psi}}_{yz}}{24} - \frac{c_{55}h\kappa\alpha_m\tilde{\overline{w}}_0}{2} = 0,\end{aligned} \tag{27}$$

$$\begin{aligned}&\left[\frac{h^3\left(c_{66}\alpha_m^2 + 2c_{22}\beta_n^2 + 2s^2\rho_p\right)}{24} + \frac{c_{44}h\kappa}{2}\right]\tilde{\overline{\psi}}_{yz}\\&\quad + \frac{h^3\alpha_m\beta_n(2c_{12}+c_{66})\tilde{\overline{\psi}}_{xz}}{24} - \frac{c_{44}h\kappa\beta_n\tilde{\overline{w}}_0}{2} = 0,\end{aligned} \tag{28}$$

$$\begin{aligned}&\frac{h\tilde{\overline{w}}_0\left(c_{55}\kappa\alpha_m^2 + c_{44}\kappa\beta_n^2 + 2s^2\rho_p\right)}{2} - \frac{c_{55}h\kappa\alpha_m\tilde{\overline{\psi}}_{xz}}{2}\\&\quad - \frac{c_{44}h\kappa\beta_n\tilde{\overline{\psi}}_{yz}}{2} = \frac{4\eta_1 t_d p_m}{\left[ab(1+st_d)\right]}.\end{aligned} \tag{29}$$

Solutions in the Fourier and Laplace transformed domains for the non-FSI case can be derived from (27)–(29) as follows:

$$\begin{aligned}\tilde{\overline{w}}_0 &= \frac{16p_m t_d\eta_1\left(\eta_3^2 - \eta_2\eta_4\right)}{\left[ab(1+st_d)\eta_6\right]},\\ \tilde{\overline{\psi}}_{xz} &= \frac{8h\kappa p_m t_d\eta_1\left(c_{44}\beta_n\eta_3 - c_{55}\alpha_m\eta_4\right)}{\left[ab(1+st_d)\eta_6\right]},\\ \tilde{\overline{\psi}}_{yz} &= \frac{8h\kappa p_m t_d\eta_1\left(c_{55}\alpha_m\eta_3 - c_{44}\beta_n\eta_2\right)}{\left[ab(1+st_d)\eta_6\right]},\end{aligned} \tag{30}$$

where

$$\begin{aligned}\eta_2 &= \frac{h\kappa c_{55}}{2} + \frac{h^3\left(2c_{11}\alpha_m^2 + c_{66}\beta_n^2 + 2s^2\rho_p\right)}{24},\\ \eta_3 &= \frac{h^3(2c_{12}+c_{66})\alpha_m\beta_n}{24},\\ \eta_4 &= \frac{h\kappa c_{44}}{2} + \frac{h^3\left(c_{66}\alpha_m^2 + 2c_{22}\beta_n^2 + 2s^2\rho_p\right)}{24},\\ \eta_5 &= \frac{h\left(\kappa c_{55}\alpha_m^2 + \kappa c_{44}\beta_n^2 + 2s^2\rho_p\right)}{2},\\ \eta_6 &= h^2\kappa^2\left[c_{44}\beta_n\left(c_{44}\beta_n\eta_2 - 2c_{55}\alpha_m\eta_3\right) + c_{55}^2\alpha_m^2\eta_4\right]\\&\quad + 4\left(\eta_3^2 - \eta_2\eta_4\right)\eta_5.\end{aligned} \tag{31}$$

4.2. Air-Backed. Considering the FSI effects for an air-backed plate case and substituting (1), (3), and (16)–(19) into (12)–(14), the fundamental (12)-(13) that are rewritten in the Fourier transformed domain are the same as (23)-(24). These rewritten equations in both the Fourier and Laplace transformed domains are the same as (27)-(28), respectively. Moreover, (14) can be rewritten as follows:

$$\frac{c_w\rho_w\dot{\overline{w}}_0 + h\rho_p\ddot{\overline{w}}_0 - c_{55}h\kappa\alpha_m\overline{\psi}_{xz}}{2} - \frac{c_{44}h\kappa\beta_n\overline{\psi}_{yz}}{2} + \frac{h\kappa\overline{w}_0\left(c_{55}\alpha_m^2 + c_{44}\beta_n^2\right)}{2} = \frac{8\eta_1 p_m \exp\left(-t/t_d\right)}{ab}. \tag{32}$$

Applying the Laplace transform to (32) yields:

$$\tilde{\overline{w}}_0\left[\frac{h\left(c_{55}\kappa\alpha_m^2 + c_{44}\kappa\beta_n^2 + 2s^2\rho_p\right)}{2} + sc_w\rho_w\right] - c_{55}h\kappa\alpha_m \times \frac{\tilde{\overline{\psi}}_{xz}}{2} - \frac{c_{44}h\kappa\beta_n\tilde{\overline{\psi}}_{yz}}{2} = \frac{8\eta_1 t_d p_m}{\left[ab\left(st_d+1\right)\right]}. \tag{33}$$

Solving (27), (28), and (33) yields the analytical solutions in the Fourier and Laplace transformed domains for an air-backed plate:

$$\begin{aligned}
\tilde{\overline{w}}_0 &= \frac{32p_m t_d\eta_1\left(\eta_3^2 - \eta_2\eta_4\right)}{\left[ab\left(1+st_d\right)\eta_8\right]},\\
\tilde{\overline{\psi}}_{xz} &= \frac{16h\kappa p_m t_d\eta_1\left(c_{44}\beta_n\eta_3 - c_{55}\alpha_m\eta_4\right)}{\left[ab\left(1+st_d\right)\eta_8\right]},\\
\tilde{\overline{\psi}}_{yz} &= \frac{16h\kappa p_m t_d\eta_1\left(c_{55}\alpha_m\eta_3 - c_{44}\beta_n\eta_2\right)}{\left[ab\left(1+st_d\right)\eta_8\right]},
\end{aligned} \tag{34}$$

where

$$\begin{aligned}
\eta_7 &= \frac{h\left(\kappa c_{55}\alpha_m^2 + \kappa c_{44}\beta_n^2 + 2s^2\rho_p\right)}{2} + sc_w\rho_w,\\
\eta_8 &= h^2\kappa^2\left[c_{44}\beta_n\left(c_{44}\beta_n\eta_2 - 2c_{55}\alpha_m\eta_3\right) + c_{55}^2\alpha_m^2\eta_4\right]\\
&\quad + 4\left(\eta_3^2 - \eta_2\eta_4\right)\eta_7.
\end{aligned} \tag{35}$$

4.3. Water-Backed. Similarly to the air-backed case, for the water-backed plate case, the fundamental (12)-(13) are rewritten in both Fourier and Laplace transformed domains and are the same as (27)-(28). Moreover, by substituting (1), (4), and (16)–(19) into (14), this equation can be rewritten as

$$\frac{2c_w\rho_w\dot{\overline{w}}_0 + h\rho_p\ddot{\overline{w}}_0 - c_{55}h\kappa\alpha_m\overline{\psi}_{xz}}{2} - \frac{c_{44}h\kappa\beta_n\overline{\psi}_{yz}}{2} + \frac{h\kappa\overline{w}_0\left(c_{55}\alpha_m^2 + c_{44}\beta_n^2\right)}{2} = \frac{8\eta_1 p_m \exp\left(-t/t_d\right)}{ab}. \tag{36}$$

Applying the Laplace transform to (36) yields:

$$\tilde{\overline{w}}_0\left[\frac{h\left(c_{55}\kappa\alpha_m^2 + c_{44}\kappa\beta_n^2 + 2s^2\rho_p\right)}{2} + 2sc_w\rho_w\right] - c_{55}h\kappa\alpha_m \times \frac{\tilde{\overline{\psi}}_{xz}}{2} - \frac{c_{44}h\kappa\beta_n\tilde{\overline{\psi}}_{yz}}{2} = \frac{8\eta_1 t_d p_m}{\left[ab\left(st_d+1\right)\right]}. \tag{37}$$

Solving (27), (28), and (37) yields the analytical solutions in the transformed domains:

$$\begin{aligned}
\tilde{\overline{w}}_0 &= \frac{32p_m t_d\eta_1\left(\eta_3^2 - \eta_2\eta_4\right)}{\left[ab\left(1+st_d\right)\eta_{10}\right]},\\
\tilde{\overline{\psi}}_{xz} &= \frac{16h\kappa p_m t_d\eta_1\left(c_{44}\beta_n\eta_3 - c_{55}\alpha_m\eta_4\right)}{\left[ab\left(1+st_d\right)\eta_{10}\right]},\\
\tilde{\overline{\psi}}_{yz} &= \frac{16h\kappa p_m t_d\eta_1\left(c_{55}\alpha_m\eta_3 - c_{44}\beta_n\eta_2\right)}{\left[ab\left(1+st_d\right)\eta_{10}\right]},
\end{aligned} \tag{38}$$

where

$$\begin{aligned}
\eta_9 &= \frac{h\left(\kappa c_{55}\alpha_m^2 + \kappa c_{44}\beta_n^2 + 2s^2\rho_p\right)}{2} + 2sc_w\rho_w,\\
\eta_{10} &= h^2\kappa^2\left[c_{44}\beta_n\left(c_{44}\beta_n\eta_2 - 2c_{55}\alpha_m\eta_3\right) + c_{55}^2\alpha_m^2\eta_4\right]\\
&\quad + 4\left(\eta_3^2 - \eta_2\eta_4\right)\eta_9.
\end{aligned} \tag{39}$$

At this point, the analytical solutions for the three cases in the transformed domains have been obtained, and subsequently, the final solutions for the displacement distributions are obtained by the inversion of Fourier and Laplace transforms. The Fourier inversion can be carried out by utilizing (16), and the Laplace inversion can be carried out by adopting an analytical inversion method, which is expressed as follows [23–25].

The solutions for the displacements in (30), (34), and (38) can be rewritten in a general and simpler form as

$$\tilde{f}(s) = \frac{A(s)}{B(s)}, \tag{40}$$

where $\tilde{f}(s)$ denotes a displacement solution in the Laplace transformed domain, both functions $A(s)$ and $B(s)$ are in form of polynomials, and the degree of $A(s)$ is less than that of $B(s)$.

If all the roots of $B(s)$ are simple, then the original function is [23–25]

$$f(t) = \sum_{k=1}^{n}\frac{A(s_k)}{B'(s_k)}\exp\left(s_k t\right). \tag{41}$$

When $B(s)$ has complex roots, the original function becomes [23–25]

$$f(t) = \sum_{i=1}^{n_r}\frac{A(s_i)}{B'(s_i)}\exp\left(s_i t\right) + 2\,\mathrm{Re}\sum_{j=1}^{n_c}\frac{A(s_j)}{B'(s_j)}\exp\left(s_j t\right), \tag{42}$$

where s_k, s_i, and s_j are the roots, the real roots, and the complex roots with positive imaginary parts of $B(s)$, respectively, n, n_r, and n_c are the numbers of s_k, s_i, and s_j, respectively, the superscript ($'$) stands for the derivative with respect to s, and Re stands for the real part of the complex number.

Table 1: The parameters of the underwater explosion, geometry for the analytical and FEA methods, and the Fourier transform parameters.

W (kg)	S (m)	p_m (MPa)	t_d (ms)	a (m)	b (m)	h (m)	ρ_w (kg/m^3)	c_w (m/s)	x_{len} (m)	y_{len} (m)	z_{len} (m)	$m_{\max}$	$n_{\max}$
1	5	8.46	0.137	1	1	0.15	1000	1500	0.025	0.025	0.025	11	11

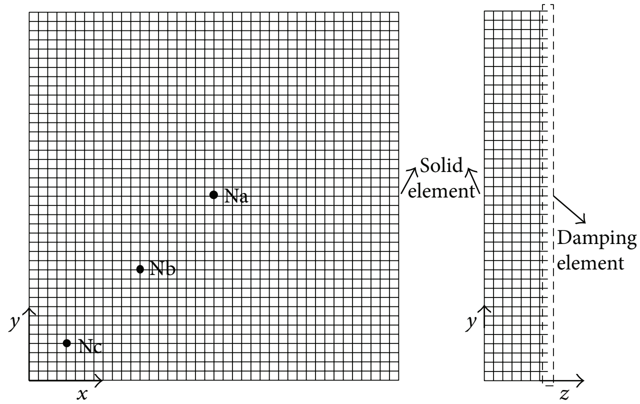


Figure 2: The model used for the FEA.

Table 2: The elastic constants of the plate material.

ρ_p (kg/m^3)	c_{11} (GPa)	c_{12} (GPa)	c_{22} (GPa)	c_{44} (GPa)	c_{55} (GPa)	c_{66} (GPa)
7850	215.54	18.48	61.58	55.26	161.54	161.54

Table 3: Geometry and elastic parameters of the aluminum plate.

Length a (m)	Width b (m)	Thickness h (m)	Young's modulus E (GPa)	Poisson's ratio μ	Density ρ (kg/m^3)
1	1	0.01	70	0.3	2700

5. Results and Discussions

A Mathematica package was developed for the analytical method described herein. First, for an air-backed plate, the results of the analytical method are validated by a comparison with a finite element analysis (FEA) performed using the commercial software ANSYS. Subsequently, the influences of the FSI and the thickness are investigated.

5.1. Validation of the Analytical Method by Comparing with FE Method. For an air-backed plate, the analytical method is validated by comparing the results of this method with those obtained by the FEA. A simply supported plate is shown in Figure 1, and the corresponding FEA model is shown in Figure 2. The plate is simulated by solid45 elements, and Combin14 damping elements are used to the damping term in (1). The elastic properties, explosive characteristics, and geometric parameters are listed in Tables 1 and 2, where x_{len} and y_{len} are the element lengths in the x and the y directions, respectively, of the FEA model for the laminated plate. The variable z_{len} is the element length in the z direction, and $m_{\max}$ and $n_{\max}$ are the maximum values of m and n, respectively.

In Figure 3, the response-time histories of $w, \sigma_{xx}, \sigma_{yy}, \sigma_{xy}, \sigma_{xz}, \sigma_{yz}$, and σ_m of the back face ($z = h/2$) and middle face ($h = 0$) at Na ($x = y = 0.5$ m), Nb ($x = y = 0.3$ m), and Nc ($x = y = 0.1$ m) are calculated by the analytical and the FEA methods for the air-backed case and are compared. In Figure 4, the distributions of the deflection w and the von Mises stress σ_m on the back face of the plate when $t = 0.45$ ms are calculated by the analytical and the FEA methods for the air-backed case and are compared. The time-histories and the distributions of the responses between the two solution methods correspond closely.

5.2. Validation of the Analytical Method by Comparing with Experiment Results. In this section, the analytical method is validated by comparing the acceleration of the air-backed aluminum plate calculated by this method with those obtained by the Hung et al.'s underwater explosion experiment result [26]. Hung performed the experiment within a $4 \times 4 \times 4$ m water tank. The isotropic aluminum target plate was fixed to an empty steel casing, known as a "shock rig", and the button of which was fixed to a steel base. In this paper, it is assumed that the plate is simply supported. The geometry and elastic properties of the plate is shown in Table 3. 1 g charge ($W = 1$ g) was located on the normal line through the center of the plate. An accelerometer was placed in the center of the plate and operated at a sampling rate of 2.0 MHz. Four tests, with different standoff distances, were carried out, although only the case with a standoff distance of 0.7 m ($S = 0.7$ m) was used for the comparisons made here. The water density is $\rho_w = 1000$ kg/m^3, and the shock wave (acoustic) velocity in the water is $c_w = 1500$ m/s.

Figure 5 compares the acceleration time histories at the center of the plate between analytical method and experiment [26]. Obviously, the time histories by the two methods agree with each other closely in the overall trend. There are some differences between the two time history curves. The acceleration peak value of analytical method is 25.6 percent greater than that of experiment. The growth can be attributed to the fact that the pressure precursor is ignored in the incident wave profile. It is understandable since the pressure precursor significantly decreases the peak of incident pressure according to [27]. In addition, there are some high-frequency vibrations in the experiment record, which are missing in the curve by the analytical method. The difference may be caused by the FSI model formulated by Taylor used here which ignores the bubble impulsive pressure. In general, the analytical method proposed in this paper and its FSI model have been validated.

5.3. The Influence of the Fluid-Solid Interaction and the Thickness. In this section, the effects of the FSI and the thickness are investigated by the analytical method. Five different thicknesses ($h = 0.05$ m, 0.10 m, 0.15 m, 0.20 m,

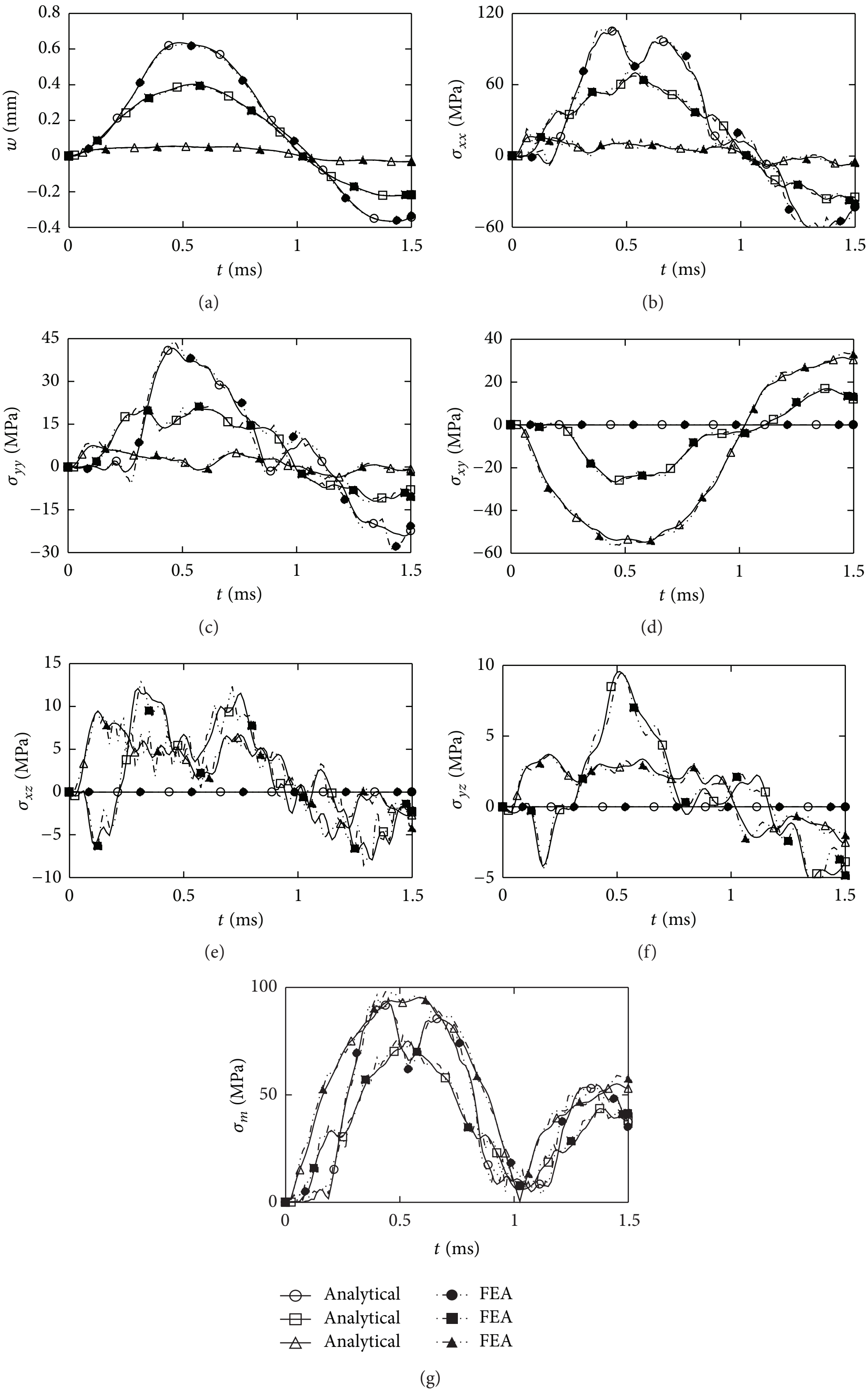


FIGURE 3: Comparisons of the response-time histories calculated by both the analytical and the FEA methods for the air-backed cases: (a) the time histories of the deflection w of the back face at Na, Nb, and Nc; (b) the time histories of the membrane normal stress σ_{xx} of the back face at Na, Nb, and Nc; (c) the time histories of the membrane normal stress σ_{yy} of the back face at Na, Nb, and Nc; (d) the time histories of the membrane shear stress σ_{xy} of the back face at Na, Nb, and Nc; (e) the time histories of the transverse shear stress σ_{xz} of the middle face at Na, Nb, and Nc; (f) the time histories of transverse shear stress σ_{yz} of the middle face at Na, Nb, and Nc; and (g) the time histories of the von Mises stress σ_m of the back face at Na, Nb, and Nc.

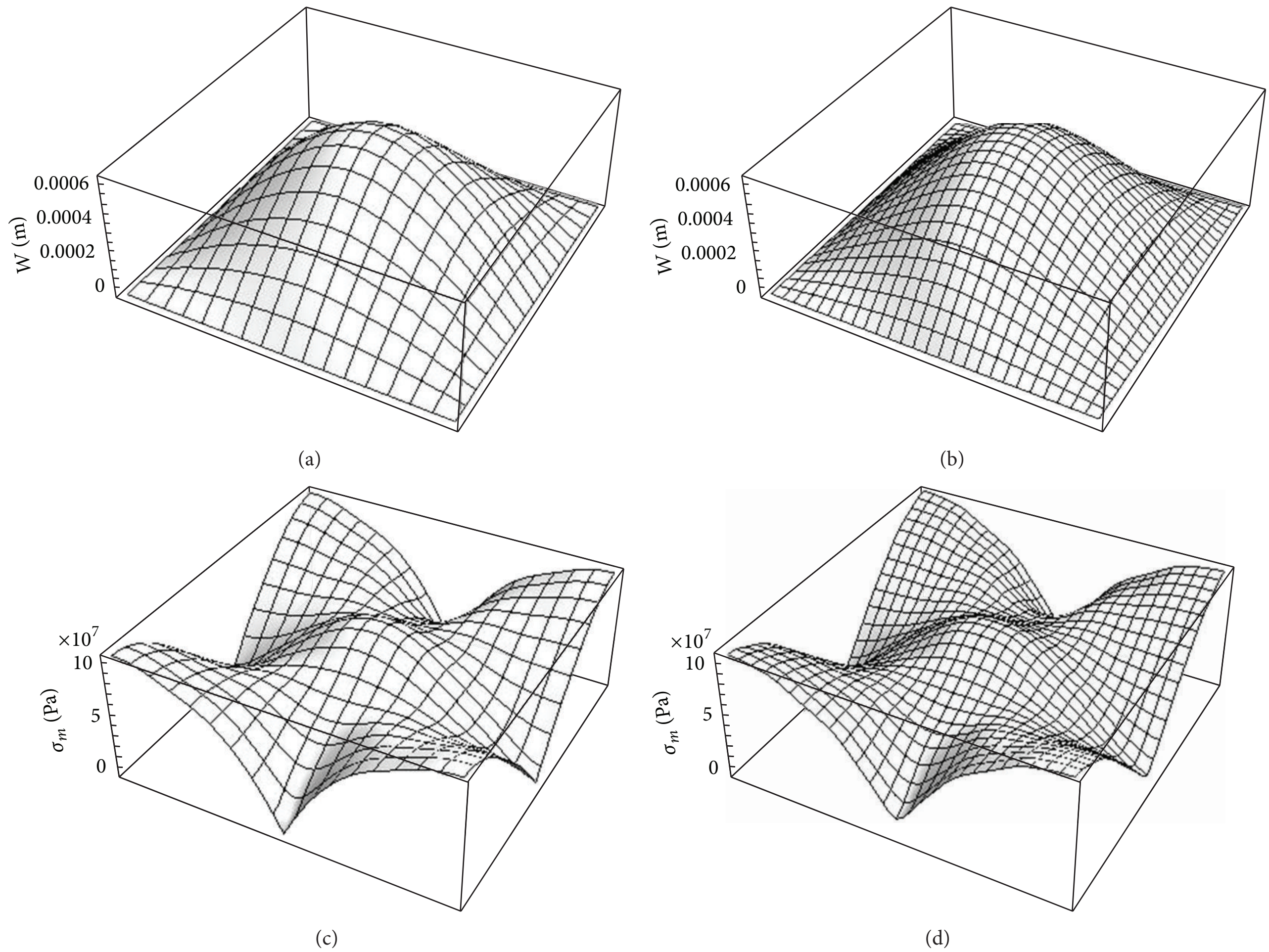


Figure 4: Comparisons of the responses of the plate calculated by both the analytical and the FEA methods for the air-backed case when $t = 0.45$ ms: (a) the deflection w of the back face of the plate calculated by the analytical method; (b) the deflection w of the back face of the plate calculated by the FEA method; (c) the von Mises stress σ_m of the back face of the plate calculated by the analytical method; and (d) the von Mises stress σ_m of the back face of the plate calculated by the FEA method.

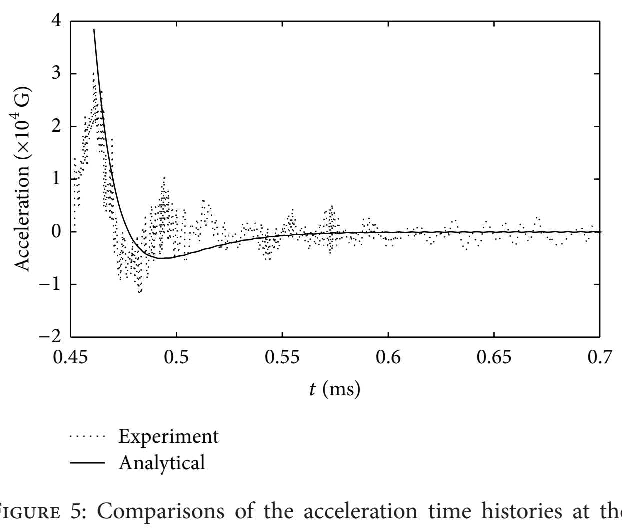


Figure 5: Comparisons of the acceleration time histories at the center of the plate between analytical method and experiment [26].

0.25 m, and 0.30 m) and three cases, including the non-FSI, air-backed, and water-backed cases, are considered. The elastic properties, explosive characteristics, and geometric parameters, except for the thickness, are listed in Tables 1 and 2. In Figure 6, the extreme values of the response-time histories of w, σ_{xx}, and σ_{xz} of the different thicknesses and the FSI cases are compared. The thickness and FSI effects significantly influence the reaction of the laminated plate.

As shown in Figures 6(a), 6(d), and 6(g), all of the extreme values of the response-time histories for the non-FSI case decrease as the thickness increases. As shown in Figures 6(b), 6(c), and 6(e), the deflection, w, and the membrane normal stress σ_{xx} in the air-backed and water-backed cases also decrease as the thickness increases. However, as shown in Figures 6(f), 6(h), and 6(i), the peak values of the response-time histories for the case of $h = 0.05$ m are less than those of $h = 0.10$ m.

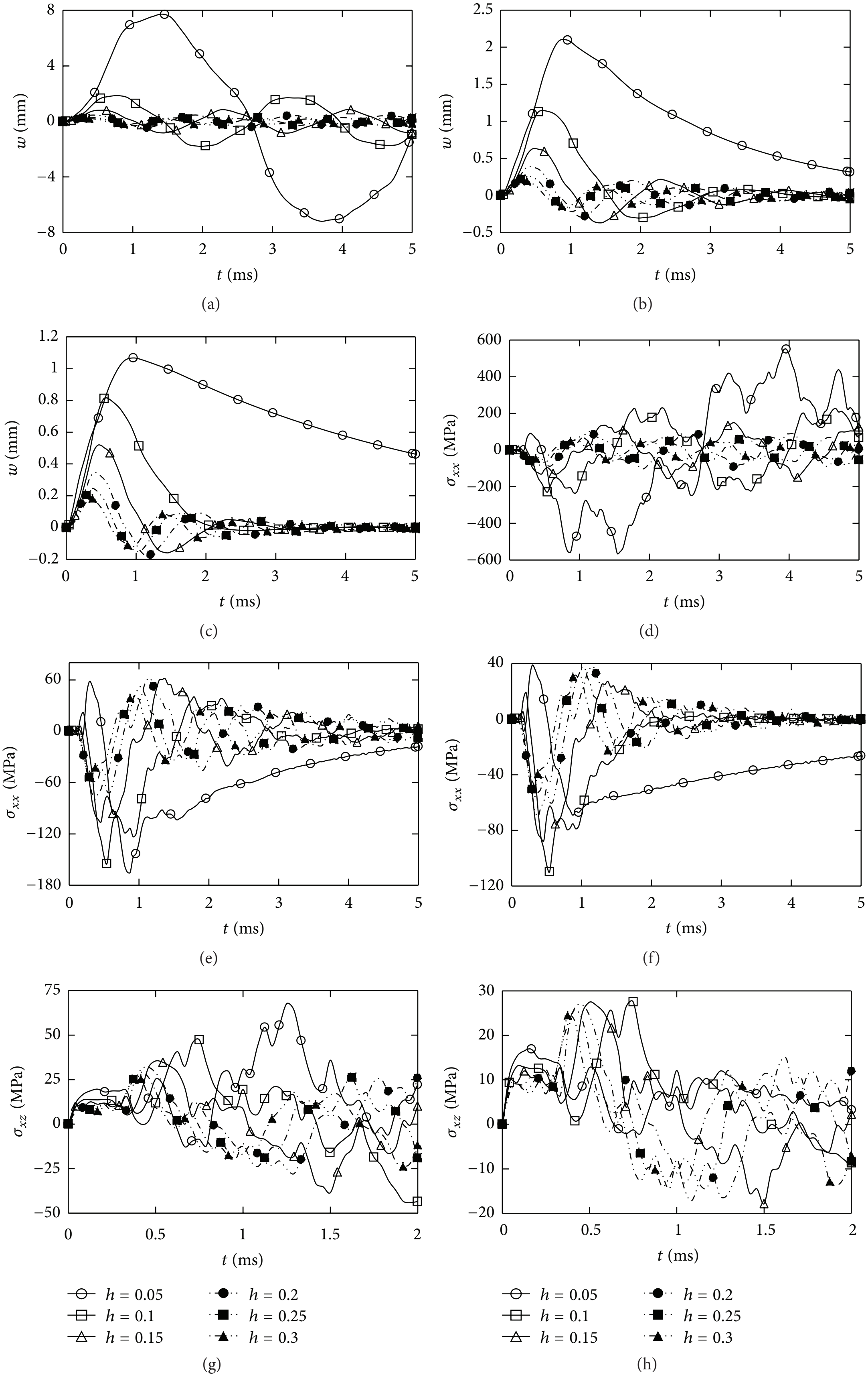
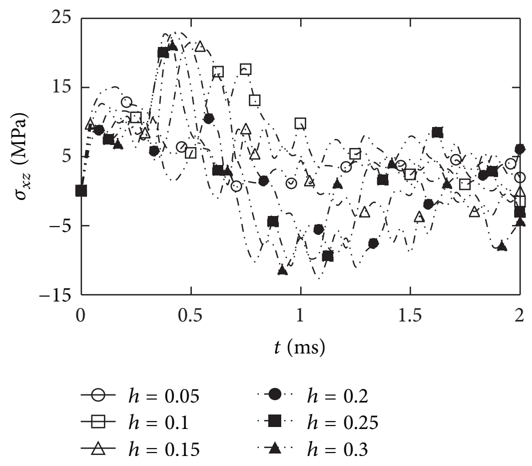


Figure 6: Continued.



(i)

FIGURE 6: Comparisons of the responses of the plate for five different thicknesses: (a) the time history of the deflection w for the plate at $x = a/2$, $y = b/2$, and $z = h/2$ for the non-FSI case; (b) the time history of the deflection w for the plate at $x = a/2$, $y = b/2$, and $z = h/2$ for the air-backed case; (c) the time history of the deflection w for the plate at $x = a/2$, $y = b/2$, and $z = h/2$ for the water-backed case; (d) the time history of the membrane normal stress σ_{xx} for the plate at $x = a/2$, $y = b/2$, and $z = h/2$ for the non-FSI case; (e) the time history of the membrane normal stress σ_{xx} for the plate at $x = a/2$, $y = b/2$, and $z = h/2$ for the air-backed case; (f) the time history of the membrane normal stress σ_{xx} for the plate at $x = a/2$, $y = b/2$, and $z = h/2$ for the water-backed case; (g) the time history of the transverse shear stress σ_{xz} for the plate at $x = 0$, $y = 0.5$, and $z = 0$ for the non-FSI case; (h) the time history of the transverse shear stress σ_{xz} for the plate at $x = 0$, $y = 0.5$, and $z = 0$ for the air-backed case; and (i) the time history of the transverse shear stress σ_{xz} for the plate at $x = 0$, $y = 0.5$, and $z = 0$ for the water-backed case.

6. Conclusions

It is in the interest of designers to know the transient state of stress in advance that develops in a plate during an underwater explosion. However, to the best of the authors' knowledge, the analytical method in the relevant literature for the dynamic response of plates to underwater explosion does not consider the FSI differences between the different positions on front and back faces of plate. The purpose of this paper is to fill this gap. Involving the planar wave assumption, an analytical method was proposed based on the FSI model formulated by Taylor [15], the Mindlin plate theory, Laplace transform, and its inversion to model the elastic dynamic response of a plate subjected to an underwater explosion. This new analytical method builds upon the Reissner-Mindlin plate theory and includes the shear deformation and the FSI (fluid-solid interaction) effect. Additionally, the new method is three-dimensional, considers the moment and the membrane stress in the plate, and examines the full profile of possibilities, such as the non-FSI, the air-backed, and water-backed cases.

The response-time histories and the response distributions on the plate in terms of displacements and stresses between the analytical method and the FEA method were compared to validate the analytical method, and good agreement was found. Subsequently, the simplified method for fluid-solid interaction as well as the analytical method are validated by comparison of the acceleration at the center of an air-backed plate between analytical method and the experiment performed by Hung et al. [26].

The thickness and FSI effects significantly influence the reaction of the laminated plate. All of the extreme values of the response-time histories for the non-FSI case decrease as the thickness increases. However, considering the FSI effects for the air-backed and water-backed cases, some peak values of the response-time histories for the $h = 0.05$ m case are less than those of $h = 0.10$ m.

The inclusion of FSI effects in designs for plates is vital for a full understanding of the dynamic response, and the results can be used as benchmark solutions in further research. In addition, several types of fluid-solid interaction effects during UNDEX near the plate are coupled in the present method, including Taylor's FSI effects, bending-stretching effects, and simply boundary effects [12]; the respective effect of each type is not presented here and will be discussed elsewhere.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the National Natural Science Foundation of China (nos. 51209185, 51079127, 51179171, and 51279180), China Postdoctoral Science Foundation (no. 2013M531462), and the Postdoctoral Science Foundation of Zhejiang Province (no. Bsh1202057). Special thanks should go to the Ansys and Mathematica software used in this work.

References

[1] R. Rajendran and J. M. Lee, "Blast loaded plates," *Marine Structures*, vol. 22, no. 2, pp. 99–127, 2009.

[2] R. Rajendran and K. Narasimhan, "Linear elastic shock response of plane plates subjected to underwater explosion,"

International Journal of Impact Engineering, vol. 25, no. 5, pp. 493–506, 2001.

[3] R. Rajendran, J. K. Paik, and B. J. Kim, "Design of warship plates against underwater explosions," *Ships and Offshore Structures*, vol. 1, no. 4, p. 347, 2006.

[4] G. M. Genin and V. Birman, "Micromechanics and structural response of functionally graded, particulate-matrix, fiber-reinforced composites," *International Journal of Solids and Structures*, vol. 46, no. 10, pp. 2136–2150, 2009.

[5] T. Hause, "Advanced functionally graded plate-type structures impacted by blast loading," *International Journal of Impact Engineering*, vol. 38, no. 5, pp. 314–321, 2011.

[6] L. Librescu, S. Y. Oh, and J. Hohe, "Dynamic response of anisotropic sandwich flat panels to underwater and in-air explosions," *International Journal of Solids and Structures*, vol. 43, no. 13, pp. 3794–3816, 2006.

[7] L. Librescu, S. Y. Oh, and J. Hohe, "Linear and non-linear dynamic response of sandwich panels to blast loading," *Composites Part B*, vol. 35, no. 6–8, pp. 673–683, 2004.

[8] L. Librescu, T. Hause, and C. J. Camarda, "Geometrically nonlinear theory of initially imperfect sandwich curved panels incorporating nonclassical effects," *AIAA Journal*, vol. 35, no. 8, pp. 1393–1403, 1997.

[9] L. Librescu, S. Y. Oh, and J. Hohe, "Implication of nonclassical effects on dynamic response of sandwich structures exposed to underwater and in-air explosions," *Journal of Ship Research*, vol. 51, no. 2, pp. 83–93, 2007.

[10] J. Leblanc and A. Shukla, "Dynamic response of curved composite panels to underwater explosive loading: experimental and computational comparisons," *Composite Structures*, vol. 93, no. 11, pp. 3072–3081, 2011.

[11] G. J. McShane, V. S. Deshpande, and N. A. Fleck, "Underwater blast response of free-standing sandwich plates with metallic lattice cores," *International Journal of Impact Engineering*, vol. 37, no. 11, pp. 1138–1149, 2010.

[12] Y. L. Young, Z. Liu, and W. Xie, "Fluid-structure and shock-bubble interaction effects during underwater explosions near composite structures," *Journal of Applied Mechanics*, vol. 76, no. 5, pp. 1–10, 2009.

[13] Z. Xue and J. W. Hutchinson, "A comparative study of impulse-resistant metal sandwich plates," *International Journal of Impact Engineering*, vol. 30, no. 10, pp. 1283–1305, 2004.

[14] A. Idesman, H. Samajder, E. Aulisa, and P. Seshaiyer, "Benchmark problems for wave propagation in elastic materials," *Computational Mechanics*, vol. 43, no. 6, pp. 797–814, 2009.

[15] G. I. Taylor, "The pressure and impulse of submarine explosion waves on plates," in *The Scientific Papers of Sir Geoffrey Ingram Taylor*, vol. 3, pp. 287–303, Cambridge, UK, 1963.

[16] Z. Liu and Y. L. Young, "Transient response of submerged plates subject to underwater shock loading: an analytical perspective," *Journal of Applied Mechanics*, vol. 75, no. 4, Article ID 044504, 5 pages, 2008.

[17] R. M. McMeeking, A. V. Spuskanyuk, M. Y. He, V. S. Deshpande, N. A. Fleck, and A. G. Evans, "An analytic model for the response to water blast of unsupported metallic sandwich panels," *International Journal of Solids and Structures*, vol. 45, no. 2, pp. 478–496, 2008.

[18] Y. Liang, A. V. Spuskanyuk, S. E. Flores et al., "The response of metallic sandwich panels to water blast," *Journal of Applied Mechanics*, vol. 74, no. 1, pp. 81–99, 2007.

[19] M. T. Tilbrook, V. S. Deshpande, and N. A. Fleck, "The impulsive response of sandwich beams: analytical and numerical investigation of regimes of behaviour," *Journal of the Mechanics and Physics of Solids*, vol. 54, no. 11, pp. 2242–2280, 2006.

[20] V. S. Deshpande and N. A. Fleck, "One-dimensional response of sandwich plates to underwater shock loading," *Journal of the Mechanics and Physics of Solids*, vol. 53, no. 11, pp. 2347–2383, 2005.

[21] R. H. Cole, *Underwater Explosions*, Princeton University Press, Princeton, NJ, USA, 1948.

[22] C. F. Hung, P. Y. Hsu, and J. J. Hwang-Fuu, "Elastic shock response of an air-backedplate to underwater explosion," *International Journal of Impact Engineering*, vol. 31, no. 2, pp. 151–168, 2005.

[23] A. H. Akbarzadeh, S. K. Hosseinizad, M. R. Eslami, and M. Sadighi, "Mechanical behaviour of functionally graded plates under static and dynamic loading," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 225, no. 2, pp. 326–333, 2011.

[24] Y. Kiani, M. Shakeri, and M. R. Eslami, "Thermoelastic free vibration and dynamic behaviour of an fgm doubly curved panel via the analytical hybrid Laplace-fourier transformation," *Acta Mechanica*, vol. 223, no. 6, pp. 1199–1218, 2012.

[25] A. H. Akbarzadeh, M. Abbasi, and M. R. Eslami, "Coupled thermoelasticity of functionally graded plates based on the third-order shear deformation theory," *Thin-Walled Structures*, vol. 53, pp. 141–155, 2012.

[26] C. F. Hung, P. Y. Hsu, and J. J. Hwang-Fuu, "Elastic shock response of an air-backed plate to underwater explosion," *International Journal of Impact Engineering*, vol. 31, no. 2, pp. 151–168, 2005.

[27] Z. Liu and Y. L. Young, "Shock-structure interaction considering pressure precursor," in *Proceedings of the 28th Symposium on Naval Hydrodynamics*, pp. 12–17, Pasadena, Calif, USA, 2010.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 653895, 11 pages
http://dx.doi.org/10.1155/2013/653895

Research Article

Multidiscipline Topology Optimization of Stiffened Plate/Shell Structures Inspired by Growth Mechanisms of Leaf Veins in Nature

Baotong Li,[1] Jun Hong,[1] Suna Yan,[1] and Zhifeng Liu[2]

[1] *State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, 28th West Xianning Road, Xi'an, Shaanxi 710049, China*
[2] *School of Mechanical and Automotive Engineering, Hefei University of Technology, 193th Tunxi Road, Hefei, Anhui 230009, China*

Correspondence should be addressed to Jun Hong; jhong@mail.xjtu.edu.cn

Received 6 February 2013; Accepted 24 April 2013

Academic Editor: Zhuming Bi

Biological structures with preeminent performance in nature endow inexhaustible inspiration for creative design in engineering. In this paper, based on the observation of the natural morphogenesis of leaf veins, we put forward a simple and practical multidiscipline topology optimization method to produce the stiffener layout for plate/shell structures. This method simulates the emergence of complex branching patterns copying the self-optimization of leaf veins which always try to grow into a configuration with global optimal performances. Unlike the conventional topology optimization methods characterized by "subtraction mode," the proposed method is based on the "addition mode," giving great potential for designers to achieve more clear stiffener layout patterns rather than vague material distributions and, consequently, saving computational resources as well as enhancing availability of design outputs. Numerical studies of both static and dynamic problems considered in this paper clearly show the suitability of the proposed method for the optimal design of stiffened plate/shell structures.

1. Introduction

To satisfy the ever-increasing demands of resource and energy savings, products with light-weight design are desirable in manufacturing industries. One common and cost-effective approach is the application of stiffened plate/shell structures. For example, load-bearing components in large-scale equipments, such as the bed of machine tool, crossbeam of crane, ship hull, and automobile body, are usually box structures filled with stiffener plates distributed horizontally and vertically. It is no doubt that the mechanical performances of the load-bearing components, including stiffness, strength, and dynamic characteristics, are dependent on the layout pattern of internal stiffener plates to a great extent; thus considerable efforts have been devoted to the optimal design of stiffer layout patterns.

Following the advancement of the numerical calculation technique, the structural topology optimization design has gotten fast development over the last two decades, offering great potential for better solutions to stiffener layout design. Luo and Gea [1, 2] developed a systematic topology optimization approach to assist the optimal stiffener design of a 3D plate/shell structure, in which both stiffener location and orientation were considered to deal with static as well as eigenvalue problems. Krog and Olhoff [3] dealt with topology optimization of externally rib-stiffened and internally stiffened honey comb and sandwich Mindlin plates by an improved homogenization approach. Bojczuk and Szteleblak [4] presented and validated a computationally efficient method to calculate the sensitivity of stiffener introduction with respect to arbitrary objective functional of displacements, strains, stresses, and resonant frequencies, which can be used to determine the optimum stiffener location. Ansola et al. [5] proposed a global optimality criteria method for the presentation of both the geometry of the shell mid-plane and the layout of surface stiffeners on the shell structure. Other approaches and contributions can be found in the literature [6–10].

Since the time when the idea of stiffener layout design has been transformed into a searching problem of optimal material distribution, the subject has been continuously developed for more and more complicated structures, such as the packaging structures [11], hull structures [12, 13], and aeronautical structures [14, 15]. The previous engineering practices in topology optimization can be deemed as a "subtraction-mode-" based procedure, because the resulting layout is extracted from the design domain by eliminating redundant material iteratively. However, such kind of subtraction mode approaches can only produce a vague material distribution; the exact information about the stiffener location and orientation is not available. Therefore, expensive computational efforts and additional postprocessing treatment are required which make the design process tedious, not effective, and time consuming.

To overcome these difficulties, it would be a meaningful way to borrow some experiences from nature. Using bio-inspired approach to improve the quality of structure design has attracted increasing attention in academic and engineering field. Actually, problems on how to define the optimal load-bearing topologies are not confined to those mechanical components in machine tool, airplane, or automobile. Elaborate structures in organs in living system also provide interesting problems such as the growth of leaf veins which enable leaves to withstand the self-weight and environmental loads. It would be very natural to presume that biological configurations are formed almost deterministically so that load-bearing functions needed in organs are created efficiently. With the inspiration of branching patterns in nature, the optimal design of stiffener layout can be interpreted as an analog to a growth process of leaf veins, in which the genes of plant leaf yield a potential capability for their tissue cells to branch and to degenerate adaptively so as to grow into a steady self-optimum structure. In this end, the emergence of stiffener layout for engineering design is definitely an "addition-mode-" based procedure that is completely opposite to the subtraction mode employed in conventional approaches.

Stiffener optimal design based on the imitation of branching patterns in nature has first been introduced by Ding and Yamazaki in 2004 [16], and since then, is extensively applied [17–19]. The main idea is inspired by the observation that natural branching patterns such as the leaf venation, root, and axis system of plants always grow in such a manner that global optimal performances can be achieved. Although Ding's method can be easily implemented, its application depends primarily on an empirical formula, in which several key parameters are selected based on designer's experience. This is not very helpful especially when searching for the best design in the case of multiobjective and multiconstraint scenarios. Questions associated with the optimization parameters have been left basically not answered. So there is a clear need to develop models that can expand and deepen our understanding of the principles, properties, and mechanisms of the emergence of branching patterns in natural systems and that can also foster simple and practical methodologies for engineering design. In this regard, the present study explores the analytical laws and their corresponding equations that underlie morphogenesis phenomena in nature so as to devise a paradigm to simulate the generation process of branching patterns for topology optimization in engineering.

The paper is organized as follows. After this introduction, we briefly present the elements of branching patterns of leaf veins and their geometrical characteristics which are pertinent to the present study. Following this description, a well-founded mathematical explanation for the adaptive growth of branching patterns is derived from the Kuhn-Tucker conditions, leading to a novel optimality criterion that can serve engineering purpose for stiffener layout design. With the suggested method, we illustrate the application in static, dynamic, and multidiscipline design problems. Finally, we conclude the paper with the main findings of the present work.

2. Parametric Characterization of Leaf Veins in Nature

2.1. Geometrical Modeling of Leaf Veins. After billions of years of evolution, nature has created complex and ingenious topologies for solving engineering problems and improving design methods. A remarkable class of such topologies is the branching configuration existed in numerous biological systems whose evolution and function have fascinated and engaged scientists and engineers for centuries. Despite the advancement of computational techniques, the process of finding analytical laws that represent the growth process of branching patterns in nature still remains uncertain. Reproduction of biological growth faces challenging problems as the experimental data can only tell us what is happening, but not why it happens, and the databases of simulation results tend to be dynamic, incomplete, redundant, and very large.

Since the general approach to modeling such kind of common growth mechanism is not expected at the present stage, it would make sense to conduct case studies for some typical prototypes, in which their geometrical configurations and physical functions are described mathematically and certain relations are discussed between them. The leaf venation treated here is a living organ that not only possesses the physiological functions, but also adapts to the complex environmental loads by evolving itself into an optimum structure with two major structural patterns: (1) medial axis pattern and (2) closed loop pattern, as shown in Figure 1.

The morphogenesis of leaf veins can be deemed as (1) a continuous growth process including sprouting, branching, and degenerating activities which can finally form a complicated dichotomous hierarchical system; (2) a continuous optimization process in which both the growth rate and growth direction are adjusted dynamically so as to make it possible for the layout of leaf venation to the dependent on the growth environment. In order to apply the growth optimality to practical engineering design, our first priority is to establish a refined leaf model with venation systems so that the biological growth procedure can be simulated in a more realistic way.

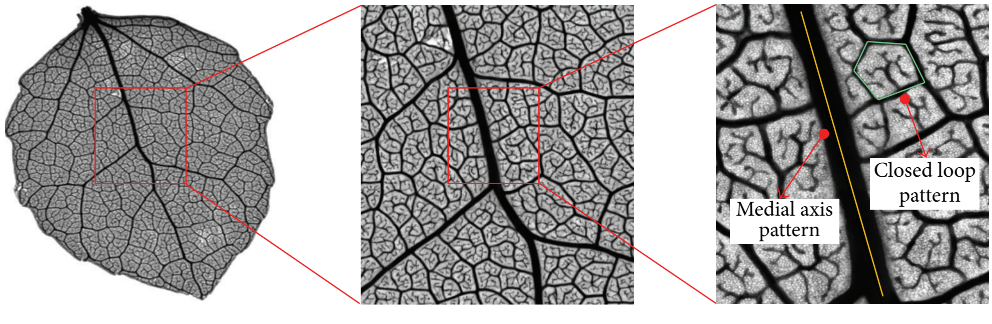


Figure 1: Leaf vein's branching patterns in different scales.

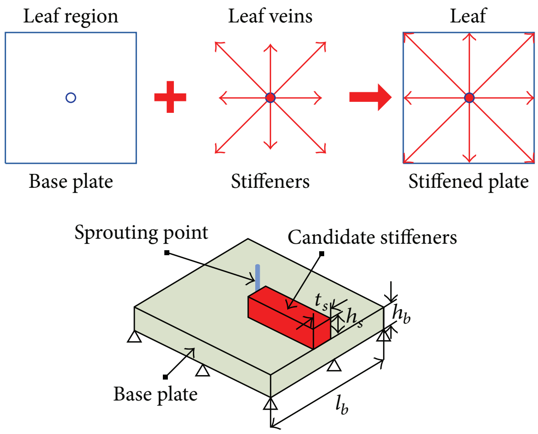


Figure 2: Geometrical modeling of leaf venation.

Although the morphologies and the growth environments of various leaf veins differ from each other, the geometrical characteristics of leaf venation can be commonly classified into two categories: one is the geometrical properties of the leaf vein itself, that is, its location, orientation, length, and thickness; the other is those of a region in the leaf blade governed by a branch, that is, its contour shape and volume. Based on this, the growth of leaf venation can be simulated as a generation process for stiffeners on a base plate as shown in Figure 2.

It can be found that the geometrical model includes both the base plate and candidate stiffeners. The base plate is utilized to model a certain region in the leaf blade governed by a new branch, the material modulus of which is defined as E_b. The plate center is assumed as the sprouting point, around which several candidate stiffeners are located symmetrically representing various possible sprouting directions of the new branch. The material modulus of the candidate stiffener is denoted as E_s. In this preliminary study, we first fix the value of E_s/E_b at 1, and the cross-sectional shape of the stiffener is supposed to be a rectangle whose width (t_s) is the same as the thickness of the base plate (h_b). Therefore, the weight of the stiffener, which will be regarded as a design variable in the simulation process, is changed only depending on the height of the stiffener (h_s).

2.2. Numerical Observation of Growth Competition. As is known, competition among candidate stiffeners during growth takes place by nonoptimal stiffeners being defeated and finally eliminated. Before the calculation continues, several assumptions are made with regard to simulating the growth competition, and they are as follows (1) an arbitrary branching point along the leaf vein is selected as a sprouting point, around which the local leaf blade is modeled as a simply supported square plate with a ratio of length to thickness of $l_b/h_b = 1000$. (2) The environmental loads are illustrated in Figure 3, where the base plate is divided into four regions, and two diagonal regions are applied with the same magnitude of the pressure. The ratio of the lower pressure to the higher pressure is defined as 0.2. (3) A certain kind of nutrition is presumed to be distributed over the base plate, which can be deemed as the distribution of the auxin carrier proteins in the mesophyll when a new leaf vein began to grow. In this model, the spatial distribution of the strain energy is simulated as the nutrient distribution, driving the new vein to sprout and branch. (4) A very small initial height ($h_{s0} = 0.01\,\text{mm}$) is introduced so as to create the seedling stiffeners, whose stiffness is a very small fraction of the whole stiffness and confirmed to have little effect on the structural performance of the base plate in the initial growth period.

It should be noted that those parameters defined above may not be accurate, yet the purpose of these assumptions is to reveal the secret design rules derived in nature. In this section, a parametric study focusing on the relationship between the height of candidate stiffeners and the plate strain energy is performed by numerical calculation based on the finite element method.

In Figure 3, the palettes of changing histories of the plate strain energy are shown as different candidate stiffeners grow from little to mature. It can be found that with the increase in the height of candidate stiffeners, the strain energy of the base plate appears as a decreasing tendency which means that stiffeners grow taller by absorbing the nutrients distributed surrounding. Then, with a further increase in the stiffener height, the average strain energy will eventually reach a convergence which implies that stiffeners are not able to grow because of lack of nutrition. That is, the growth process will stop when the strain energy of the base plate cannot be decreased anymore. Table 1 lists the quantitative relationship between h_s and R for each candidate stiffener, where h_s refers

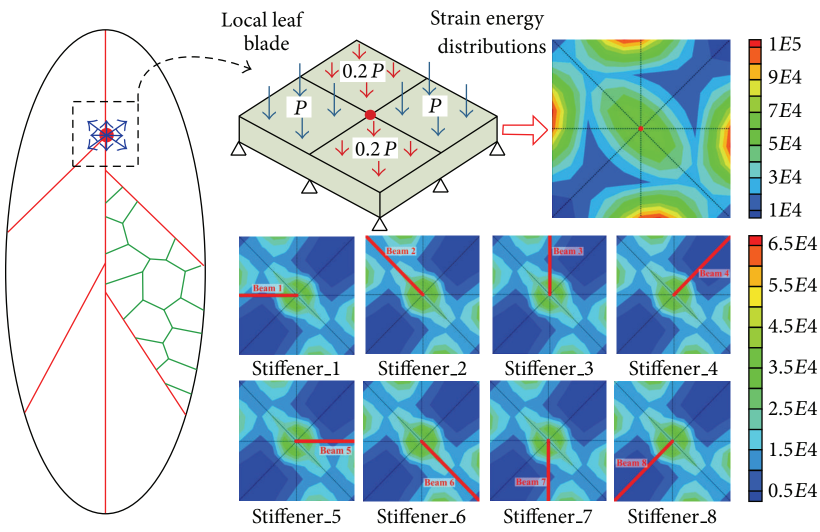


FIGURE 3: Growth competition among candidate stiffeners.

TABLE 1: Convergence values of h_s and R for each stiffener.

	B-1	B-2	B-3	B-4	B-5	B-6	B-7	B-8
h_s (mm)	7.460	9.700	7.460	9.920	7.460	9.700	7.460	9.920
R (%)	31.74	32.03	31.74	33.16	31.74	32.03	31.74	33.16

to the convergence value of the stiffener height, and R refers to the ratio of strain energy between the initial and convergence states.

From this table we can approximate where the minimum strain energy will occur as the height of the stiffener increases with the step size. An interesting phenomenon is that although the increase in each stiffener height can eventually lead to a convergence of the strain energy, these stiffeners differ from each other in the convergence values. For instance, when the height of the stiffener_1 increases to 7.46 mm, the strain energy ratio will converge to 31.74%. However, the convergence values of the stiffener height and strain energy ratio are 9.70, 32.03% and 9.92, 33.16% for the stiffener_2 and the stiffener_4, respectively. Such kind of difference reflects the growth competition among those candidate stiffeners. A reasonable view is that there may exist an ideal "balanced point," at which the weights of the stiffeners follow a distribution in such a way that the average strain energy will be minimized. The more materials the stiffeners obtain, the more opportunities they will get to win in the fierce growth competition, and the final winner will represent the sprouting direction of the new vein.

3. Mathematical Insights into Growth Competition

In order to capture some of the most essential features of the growth competition in leaf venation, an optimization concept termed guide-weight criterion [20, 21] is introduced in this section, by means of which the relationship between the growth objective and the corresponding resource constraint will be clarified to serve engineering purpose in stiffened plate/shell structure design.

3.1. Modeling of the "Balanced Point". Actually, there are two ways for a candidate stiffener to survive the competition for best. One is to reduce the amount of material used while maintaining the same mechanical performance; the other is to maintain the amount of the material used but with improved mechanical performance. To transform the implicit "balanced point" into a group of explicit mathematical equations, the Kuhn-Tucker conditions are employed to describe the growth competition problem, shown as follows:

$$\begin{aligned}
&\text{find } \mathbf{W} = [W_1, W_2, \ldots, W_N]^T \in R^N, \\
&\text{min: } F(\mathbf{W}) = \sum_{j=1}^{J} r_j \cdot f_j(\mathbf{W}), \\
&\text{subject to } G(\mathbf{W}) \le W_g, \\
&W_i^U > W_i > W_i^L, \quad i = 1, 2, \ldots, N,
\end{aligned} \tag{1}$$

where $\mathbf{W} = [W_1, W_2, \ldots, W_N]^T$ is an N-dimensional design vector, and W_i is the weight of the ith candidate stiffener; $F(\mathbf{W})$ is the weighted growth objective function, $f_j(\mathbf{W})$ is the jth subobjective function which denotes the mechanical performance such as the strain energy, strength, and natural frequency, r_j is the corresponding weight factor; $G(\mathbf{W})$ is the resource constraint function which denotes the amount of the material allocated to candidate stiffeners, and W_g is the upper limit of resource constraint.

The Lagrange function is constructed as follows:

$$L(\mathbf{W}) = F(\mathbf{W}) + \lambda \cdot \left(G(\mathbf{W}) - W_g\right), \quad (2)$$

where λ is the Lagrange multiplier.

Based on the Kuhn-Tucker conditions, at the optimal solution $\mathbf{W}^*$, the following formulas must be satisfied:

(C-1)

$$\nabla L = \nabla F(\mathbf{W}) + \lambda \cdot \nabla\left[G(\mathbf{W}) - W_g\right] \begin{cases} \leq 0, & W_i = W_i^U, \\ = 0, & W_i^U > W_i > W_i^L, \\ \geq 0, & W_i = W_i^L, \end{cases} \quad (3)$$

(C-2)

$$\lambda \geq 0, \quad (4)$$

(C-3)

$$\lambda \cdot \left[G(\mathbf{W}) - W_g\right] = 0, \quad \begin{cases} \lambda = 0, & G(\mathbf{W}) - W_g < 0, \\ \lambda \geq 0, & G(\mathbf{W}) - W_g = 0, \end{cases} \quad (5)$$

(C-4)

$$G(\mathbf{W}) \leq W_g, \quad (6)$$

(C-5)

$$W_i^U > W_i > W_i^L, \quad i = 1, 2, \ldots, N. \quad (7)$$

Multiplying W_i by (3) leads to

$$W_i \cdot \frac{\partial F}{\partial W_i} + \lambda \cdot W_i \cdot \frac{\partial G}{\partial W_i} = 0, \quad (8)$$

because

$$\frac{\partial G}{\partial W_i} = \frac{\partial\left(\sum_{i=1}^{N} W_i\right)}{\partial W_i} = 1. \quad (9)$$

When applying (9), (8) can be rewritten as

$$W_i \cdot \frac{\partial F}{\partial W_i} + \lambda \cdot W_i = 0. \quad (10)$$

Hence, the weight of the ith candidate stiffener can be expressed as

$$\begin{aligned} W_i &= -\frac{1}{\lambda} \cdot \left(W_i \cdot \frac{\partial F}{\partial W_i}\right) \\ &= -\frac{1}{\lambda} \cdot \left[W_i \cdot \left(\sum_{j=1}^{J} r_j \cdot \frac{\partial f_j}{\partial W_i}\right)\right]. \end{aligned} \quad (11)$$

The total weight of the stiffeners can be gained from the following equation:

$$W_{\text{sum}} = \sum_{i=1}^{N} W_i = -\frac{1}{\lambda} \cdot \sum_{i=1}^{N} \left[W_i \cdot \left(\sum_{j=1}^{J} r_j \cdot \frac{\partial f_j}{\partial W_i}\right)\right]. \quad (12)$$

Let

$$D_i = -W_i \cdot \left(\sum_{j=1}^{J} r_j \cdot \frac{\partial f_j}{\partial W_i}\right), \quad (13)$$

$$D_{\text{sum}} = \sum_{i=1}^{N} \left[-W_i \cdot \left(\sum_{j=1}^{J} r_j \cdot \frac{\partial f_j}{\partial W_i}\right)\right], \quad (14)$$

where D_i is the generalized weight of the ith candidate stiffener which is the negative product of the stiffener weight and the sensitivity of the growth objective function with respect to the stiffener weight, and D_{sum} is the generalized total weight of all the candidate stiffeners.

Thus, the weight of the ith candidate stiffener can be expressed by a simplified form:

$$W_i = \left(D_i \cdot \frac{W_{\text{sum}}}{D_{\text{sum}}}\right), \quad (i = 1, \ldots, N). \quad (15)$$

When iterating, it can be written as

$$W_i^{(k+1)} = \left(D_i \cdot \frac{W_{\text{sum}}}{D_{\text{sum}}}\right)^{(k)}, \quad (i = 1, \ldots, N). \quad (16)$$

The above equation is the so-called guide-weight criterion that yields an opportunity for designers to get the ideal "balanced point" in terms of weight distribution of candidate stiffeners.

In order to guarantee the convergence of the calculation, a relaxation factor α is introduced, and the iterative forms can be rewritten as

$$W_i^{(k+1)} = \alpha^{(k)} \cdot \left(D_i \cdot \frac{W_{\text{sum}}}{D_{\text{sum}}}\right)^{(k)} + \left(1 - \alpha^{(k)}\right) \cdot W_i^{(k)}, \quad (i = 1, \ldots, N). \quad (17)$$

From (17), we establish the governing criterion to allocate the growth resources to candidate stiffeners, guiding the growth process to make the weight distribution satisfy the previous Kuhn-Tucker conditions. Based on the observation and derivations above, the "balanced point" among candidate stiffeners can be assumed to take the following concise form:

$$\frac{D_1}{W_1} = \frac{D_2}{W_2} = \cdots \frac{D_i}{W_i} = \cdots = \frac{D_{\text{sum}}}{W_{\text{sum}}} = \lambda, \quad (i = 1, \ldots, N). \quad (18)$$

Equation (18) shows that in an optimal growth result, the material allocated to candidate stiffeners should be exactly proportional to the value of their generalized weight, and the "balanced point" is the Lagrange multiplier λ. That is, on the premise of unchanged growth resources, the growth result is optimal when the positive change of the objective function caused by the weight increase of any one of the involved candidate stiffeners is identical to the negative change caused by the weight decrease of any other candidate stiffeners.

The above-mentioned "balanced point" can be evaluated quantitatively by the following equation:

$$\sigma = \sqrt{\sum_{i=1}^{N} \left(\frac{W_i/W_{\text{sum}}}{D_i/D_{\text{sum}}} - 1 \right)^2}, \quad (i = 1, \dots, N), \tag{19}$$

where σ is called the "balanced degree" which should be zero for an optimal structure.

In this paper, the "balanced degree" is utilized to determine the step factor $\alpha^{(k)}$:

$$\alpha^{(k)} = \ln \frac{e^3}{N} \cdot \ln \left(\sigma^{(k)} - 0.1\sigma^{(0)} + 1 \right), \tag{20}$$

where $\sigma^{(0)}$ is the initial value of the "balanced degree," $\sigma^{(k)}$ is the "balanced degree" after kth iteration, and N is the total number of the candidate stiffeners. The iteration process would be terminated when the value of $\sigma^{(k)}$ is decreased to be one-tenth of its initial value $\sigma^{(0)}$.

3.2. Mathematical Explanation of the "Balanced Point". The findings in the previous section motivate us to explore an explanation, where the mathematical nature about the relationship among the Lagrange multiplier λ, objective function $F(X)$, and constraint boundary W_g will be clarified, so that the natural growth mechanism could be fully employed to engineering design.

By taking the first derivative of (3) with respect to the constraint boundary W_g, we have the following relation:

$$\frac{\partial L}{\partial W_g} = \frac{\partial F}{\partial W_g} + \frac{\partial \lambda}{\partial W_g} \left(G - W_g \right) + \lambda \cdot \frac{\partial G}{\partial W_g} - \lambda. \tag{21}$$

Let

$$\nabla_{W_g} W = \left[\frac{\partial W_1}{\partial W_g}, \frac{\partial W_2}{\partial W_g}, \dots, \frac{\partial W_N}{\partial W_g} \right]^T. \tag{22}$$

Equation (21) can be rewritten as

$$\begin{aligned} \frac{\partial L}{\partial W_g} &= \nabla_{W_g} W \cdot \left(\frac{\partial F}{\partial W} + \lambda \cdot \frac{\partial G}{\partial W} \right) \\ &\quad + \frac{\partial \lambda}{\partial W_g} \left(G - W_g \right) - \lambda, \end{aligned} \tag{23}$$

since:

$$\frac{\partial F}{\partial W} + \lambda \cdot \frac{\partial G}{\partial W} = 0. \tag{24}$$

Equation (21) can be further simplified as

$$\frac{\partial L}{\partial W_g} = \frac{\partial \lambda}{\partial W_g} \left(G - W_g \right) - \lambda. \tag{25}$$

According to (5), we obtain

$$\frac{\partial \lambda}{\partial W_g} \left(G - W_g \right) = 0. \tag{26}$$

Thus

$$\frac{\partial L}{\partial W_g} = -\lambda. \tag{27}$$

Consider that the optimal solution W^* satisfies the following equation:

$$L \left(W^* \right) = F \left(W^* \right). \tag{28}$$

Thus, the relationship among the Lagrange multiplier λ, objective function $F(X)$, and constraint boundary W_g can be expressed by

$$\frac{\partial F}{\partial W_g} = -\lambda. \tag{29}$$

The above equation provides very important insights into the essential features of the growth mechanism which is most efficient in carrying biological structures, such as leaf veins. Note that the negative Lagrange multiplier is equal to the sensitivity of $F(X)$ with respect to W_g; hence there are three cases, that is, $\lambda > 0$, $\lambda = 0$, and $\lambda < 0$, during the iterative growth process.

(1) When $\lambda > 0$, according to (18), one can find that $D_i > 0$, that is, $\partial F/\partial W_i < 0$, which means the weight of the ith candidate stiffener W_i should be increased with larger D_i so as to get an optimal growth result. That is, the objective function $F(X)$ would be improved with an increase in constraint boundary W_g.

(2) When $\lambda < 0$, it can be seen that the constraint boundary W_g should be decreased in order to improve the objective function $F(X)$. That means that the total weight of all the candidate stiffeners W_{sum} should be decreased with larger λ after each iterative growth step until $\lambda = 0$, so that the Kuhn-Tucker conditions would be satisfied.

(3) When $\lambda = 0$, whether for an increase or for a decrease in the constraint boundary W_g, the objective function $F(X)$ cannot be improved anymore.

4. Biologically Inspired Topology Optimization Algorithm

The preceding modeling and computational efforts focus primarily on the derivation of the optimality criterion that governs the growth behavior of branching patterns in leaf veins and have not yet given an evolutionary programming for the optimal design of natural and engineering systems. This section will elaborate how to employ such optimality criterion to reproduce the growth process of branching patterns in leaf veins and how to solve the multidiscipline topology optimization problem for stiffened plate/shell structures.

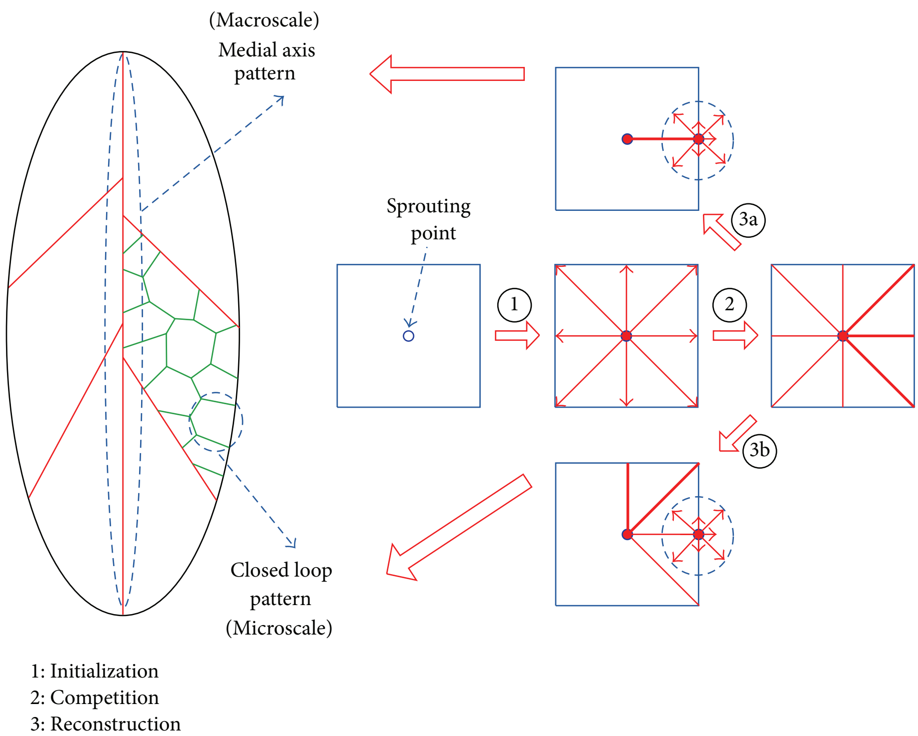


Figure 4: Simulation strategy for reproducing the adaptive growth process.

4.1. Simulation Strategy. The objective of this problem is to seek the optimal structural performance by growing a given amount of material as stiffeners. The reproduction of adaptive growth process is treated by means of a comprehensive iterative three-step approach, as shown in Figure 4.

Step 1 (initialization). The sprouting points for candidate stiffeners to grow are initialized and are included in a branching point set $\{B\}$; stiffeners extended from the sprouting points are gathered in the corresponding stiffener set $\{C\}$.

Step 2 (competition). The competition for growth resources among candidate stiffeners is simulated by performing one or more material allocation substeps, in which the growth rate of each stiffener is controlled based on the weight distribution optimized by the iterative criterion derived in Section 3 (i.e., (17)). The height of stiffener cross section (h_s) is utilized as a decision variable in the optimization process. As a result, some of the stiffeners in the set $\{C\}$ will become taller and taller, while the other stiffeners will not. Since the material allocation is treated as an inequality constrained optimization problem, the total increment of stiffener weight obtained in each iterative step, ΔW_i ($i = 1, 2, \ldots, s$), is not greater than the given constraint boundary W_g.

Step 3 (reconstruction). During the growth process, the stiffener layout is continuously reconstructed by obeying the branching and degenerating rules like those for leaf veins. For example, when a stiffener grows up to a height that is greater than the specified upper limit h_b, a subtle control is activated endowing this stiffener with capabilities for branching operation, in which two ends of the stiffener are added into the set $\{B\}$, and new stiffeners around the new sprouting points are also supplemented into the set $\{C\}$. Likewise, if the height of candidate stiffeners is less than the specified lower limit h_d, the degenerating operation will be activated, where the involved stiffeners are eliminated from the set $\{C\}$ and their ends are also deleted from the set $\{B\}$. By doing this, the best potential direction can be selected for stiffeners to grow in the next iterative step.

Both competition and reconstruction loops are repeated alternatively until there are no any material resources available from the specified growth environment. That is, the growth process will stop when the weight increment of the candidate stiffeners reaches the given upper limit W_p. The difference of the growth process between the macroscale pattern (i.e., the medial axis pattern) and the microscale pattern (i.e., the closed loop pattern) is mainly reflected in the final step of the simulation process (i.e., the reconstruction step). For the medial axis pattern, the only winner in each growth step is just the tallest stiffener that would be maintained for branching, while others are eliminated. However, for the closed loop pattern, stiffeners whose heights are greater than the specified threshold value would be eligible to participate in the next round of competition, thus yielding opportunity to form a closed loop in the following growth process.

4.2. Numerical Application. To validate the effectiveness of the reported bio-inspired methodology, three numerical examples are presented by generating the optimal stiffener

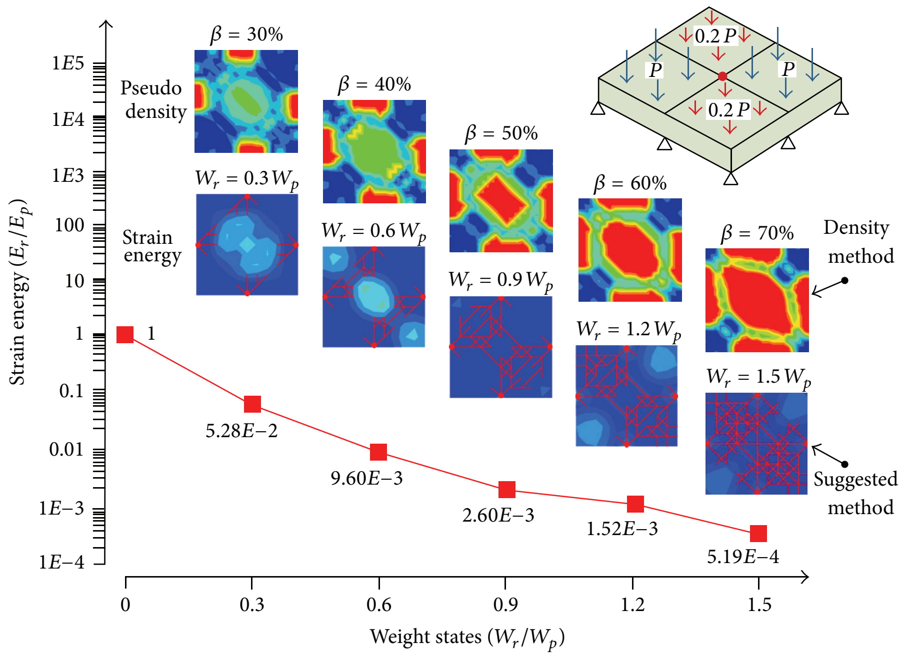


Figure 5: Comparison of stiffener layouts produced by the suggested method and the density method for static design problem.

layouts for a simply supported square plate with nonuniformly distributed pressure shown in Figure 3. The first example is a static design problem pursuing the stiffest stiffener layout so as to minimize its average strain energy; the second example is a dynamic design problem that aims to maximize the natural frequency of its first mode; and the third example is a multidiscipline design problem in which the importance of each objective is determined by putting in appropriate weights (i.e., $\omega_{\text{Static}} = 0.6$ and $\omega_{\text{Dynamic}} = 0.4$). Before getting into the actual design, the sprouting points are specified firstly on the four central points of the plate edges. The total increment of stiffener weight, W_r, is specified as $1.5W_p$, where W_p denotes the initial weight of the base plate. The upper limit for stiffeners to branch is set as $h_b = 10^4 h_d$, where h_d is the lower limit for stiffeners to degenerate. Based on these assumptions, the branch-like stiffener layouts with the weight fraction varied from 0.3 to 1.5 are constructed. To facilitate comparison, a conventional algorism based on the density method is also applied to search for the optimal material distribution, in which the ratio of retained material β is controlled from 30% to 70%, as shown in Figures 5, 6, 7, and 8.

It can be found that the suggested method is able to produce the effective stiffener layouts adaptively to various design objectives. For the static design problem, the strain energy of the final stiffened structure deceases nearly 1927 times compared to the base plate without stiffener; for the dynamic design problem, the first eigenvalue of the final stiffened structure increases nearly 12 times compared to the base plate without stiffener; and for the multidiscipline design problem, the strain energy of the final stiffened structure decreases nearly 325 times with the first eigenvalue increased about 5 times.

It can also be found that the simulated branching patterns have both higher and lower height levels, which can be interpreted to be responsible for the load-bearing properties of leaf veins in nature that the higher branches (i.e., the central veins) bear most of the principal stress, and the lower branches (i.e., the lateral veins) are the tracing of the gradient of shear stress. Obviously, the generated stiffeners are primarily concentrated in the regions which are in qualitative consistence with the optimal material distributions resulted from the conventional algorism. However, the complex branching patterns produced by the suggested method give rise to flexibility and distinctiveness of design outputs. The flexibility here means that the bio-inspired method that could generate optimal load-bearing topologies would also possess a higher tolerance to variations of design objectives which is an unintended benefit for survival in a harsh growth environment. Design results produced by the presented method are relatively clear stiffener distribution patterns rather than vague material distributions, which will help to minimize effort, time and cost during the product development.

5. Conclusions

In this work, we introduced a bio-inspired topology optimization method that integrates the morphogenesis law of leaf veins with conditioned extremum optimization theory for modeling, generating, and evolving complex branching patterns so as to identify the optimal load-bearing topology of stiffened plate/shell structures. Based on the results presented in this paper, the following conclusions can be drawn.

(1) The reported bio-inspired methodology is a valid topology optimizer by which the design solutions

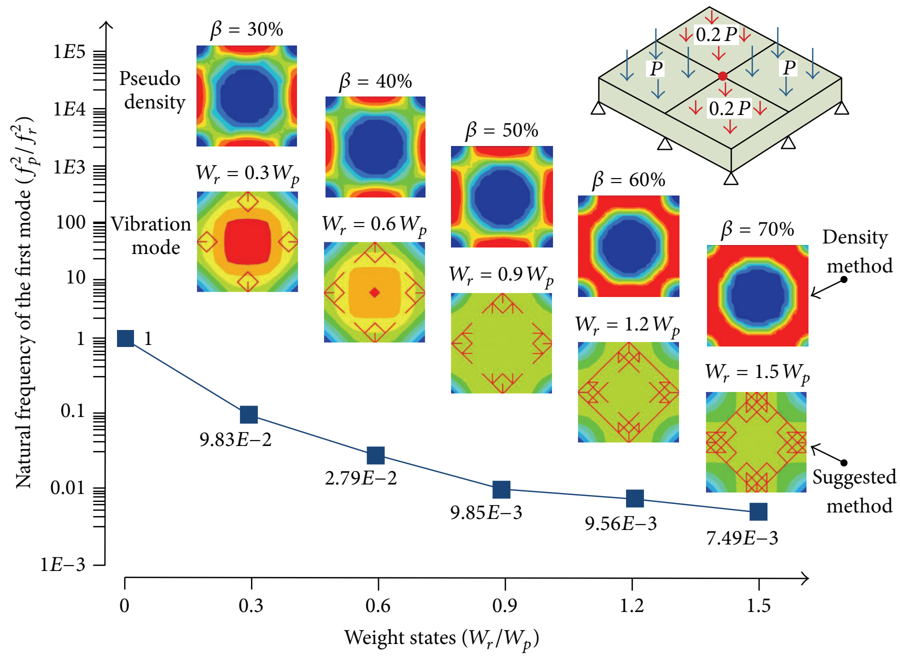


FIGURE 6: Comparison of stiffener layouts produced by the suggested method and the density method for dynamic design problem.

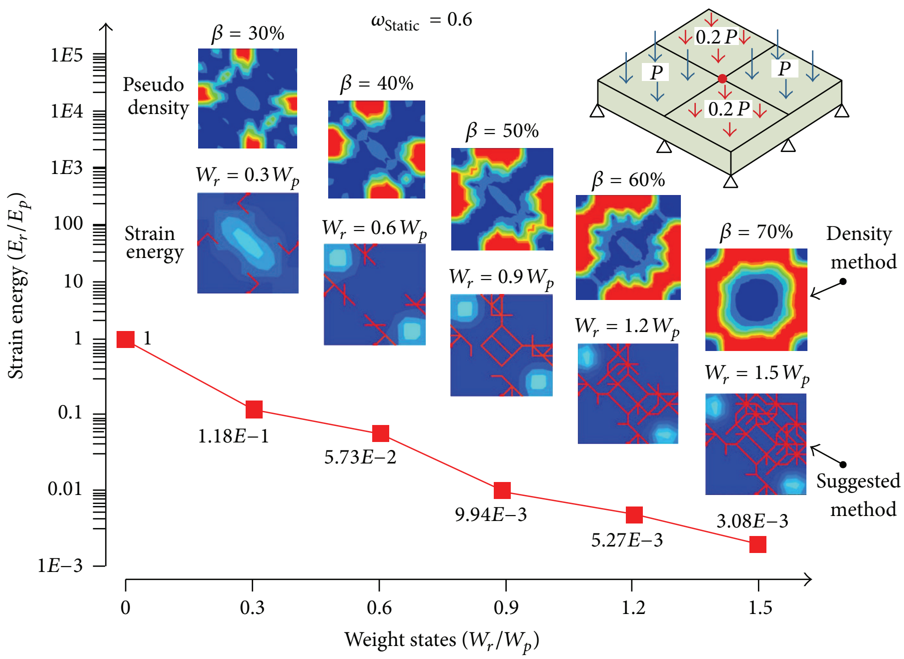


FIGURE 7: Comparison of stiffener layouts produced by the suggested method and the density method for multidiscipline design problem (strain energy).

produced are robust and distinct stiffener layout patterns rather than vague material distributions, avoiding the difficulties caused by conventional methods that normally require an additional postprocessing treatment to distinguish the real layout solution. Consequently, design efficiency is substantially improved which will help to minimize effort, time, and cost during the product development.

(2) The present work yields very important insights into the theoretical implications of natural growth mechanism for practical engineering design. Among the key problems currently facing biologists and engineers is how to understand the self-optimization characteristics of branching patterns in nature, such as the geometrical form and physical function causality in the evolution of leaf veins. The proposed simulation

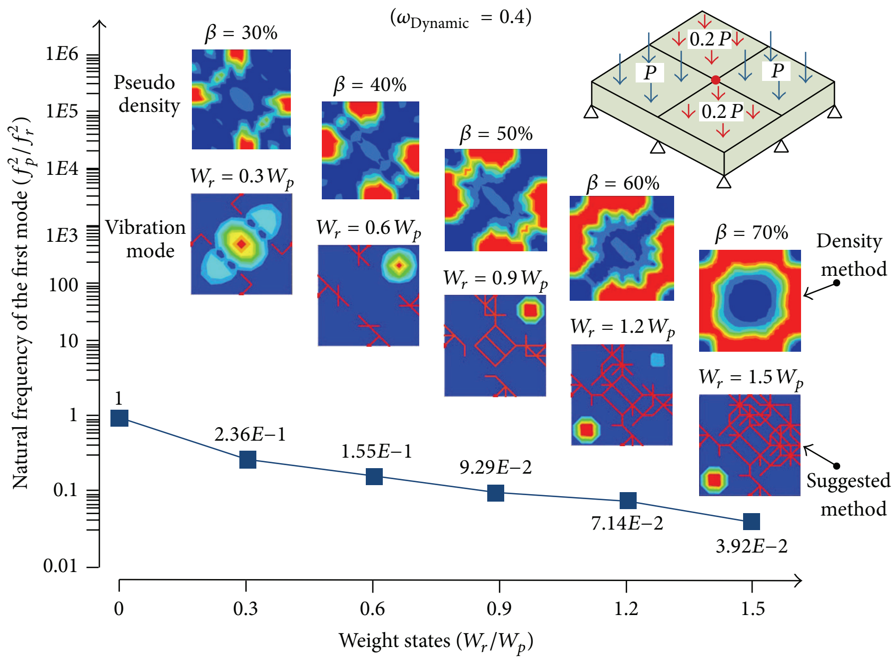


Figure 8: Comparison of stiffener layouts produced by the suggested method and the density method for multidiscipline design problem (vibration mode).

strategy can be used to verify postulations in a sufficiently controlled environment, where it is likely to isolate various physical mechanisms of interest.

The work presented in this paper is by no means complete. Future studies will include continuous model refinement. For instance, experimental and numerical studies of the relationship between typical distribution patterns of leaf veins and their mechanics self-adaptability will be carried out. The influence of different growth objectives and resource constraints will be thoroughly investigated and correlated with various natural conditions. The simulation of biological growth is a very complex task due to the strong randomness and nonlinear factors associated with the environmental conditions. In order to capture the essential features of biological growth in nature, a robust design optimization model is currently under development so that both the optimality and robustness of the simulated growth results would be taken into account in the future research.

Acknowledgments

The work reported in this paper is supported by the National Natural Science Foundation of China (51135004), the Hi-Tech Research and Development Program of China (2012AA040701), and the Major Project of High-End CNC Machine Tool and Basic Manufacturing Equipment of China (2011ZX04016-031).

References

[1] J. Luo and H. C. Gea, "A systematic topology optimization approach for optimal stiffener design," *Structural and Multidisciplinary Optimization*, vol. 16, no. 4, pp. 280–288, 1998.

[2] J. H. Luo and H. C. Gea, "Optimal bead orientation of 3D shell/plate structures," *Finite Elements in Analysis and Design*, vol. 31, no. 1, pp. 55–71, 1998.

[3] L. A. Krog and N. Olhoff, "Optimum topology and reinforcement design of disk and plate structures with multiple stiffness and eigenfrequency objectives," *Computers and Structures*, vol. 72, no. 4, pp. 535–563, 1999.

[4] D. Bojczuk and W. Szteleblak, "Optimization of layout and shape of stiffeners in 2D structures," *Computers and Structures*, vol. 86, no. 13-14, pp. 1436–1446, 2008.

[5] R. Ansola, J. Canales, J. A. Tarrago, and J. Rasmussen, "Combined shape and reinforcement layout optimization of shell structures," *Structural and Multidisciplinary Optimization*, vol. 27, no. 4, pp. 219–227, 2004.

[6] H. C. Cheng, N. Kikuchi, and Z. D. Ma, "An improved approach for determining the optimal orientation of orthotropic material," *Structural Optimization*, vol. 8, no. 2-3, pp. 101–112, 1994.

[7] S. Sakata, F. Ashida, and M. Zako, "A layout optimization of beam reinforcements to decrease thermal deformation of beam structures," *Journal of Thermal Stresses*, vol. 25, no. 11, pp. 1081–1101, 2002.

[8] Y. C. Lam and S. Santhikumar, "Automated rib location and optimization for plate structures," *Structural and Multidisciplinary Optimization*, vol. 25, no. 1, pp. 35–45, 2003.

[9] J. H. Lee, G. H. Kim, and Y. S. Park, "A geometry constraint handling technique for stiffener layout optimization problem," *Journal of Sound and Vibration*, vol. 285, no. 1-2, pp. 101–120, 2005.

[10] Y. C. Liu and Q. K. Wang, "Computational study of strengthening effects of stiffeners on regular and arbitrarily stiffened plates," *Thin-Walled Structures*, vol. 59, no. 1, pp. 78–86, 2012.

[11] C. J. Chang, B. Zheng, and H. C. Gea, "Automated design of thin-walled packaging structures," *Structural and Multidisciplinary Optimization*, vol. 35, no. 6, pp. 601–608, 2008.

[12] Y. Ha, W. Kim, and S. Cho, "Design sensitivity analysis and topology optimization method applied to stiffener layout in hull structures," *Journal of Ship Research*, vol. 50, no. 3, pp. 222–230, 2006.

[13] M. Rais-Rohani and J. Lokits, "Reinforcement layout and sizing optimization of composite submarine sail structures," *Structural and Multidisciplinary Optimization*, vol. 34, no. 1, pp. 75–90, 2007.

[14] K. Maute and M. Allen, "Conceptual design of aeroelastic structures by topology optimization," *Structural and Multidisciplinary Optimization*, vol. 27, no. 1-2, pp. 27–42, 2004.

[15] Q. Wang, Z. Z. Liu, and H. C. Gea, "New topology optimization method for wing leading-edge rib," *Journal of Aircraft*, vol. 48, no. 5, pp. 1741–1748, 2011.

[16] X. Ding and K. Yamazaki, "Stiffener layout design for plate structures by growing and branching tree model (application to vibration-proof design)," *Structural and Multidisciplinary Optimization*, vol. 26, no. 1-2, pp. 99–110, 2004.

[17] X. Ding and K. Yamazaki, "Constructal design of cooling channel in heat transfer system by utilizing optimality of branch systems in nature," *Journal of Heat Transfer*, vol. 129, no. 3, pp. 245–255, 2007.

[18] M. H. Kobayashi, "On a biologically inspired topology optimization method," *Communications in Nonlinear Science and Numerical Simulation*, vol. 15, no. 3, pp. 787–802, 2010.

[19] K. Xue and Y. X. Li, "A bionic venation growth algorithm for a stiffener layout in plate/shell structure," *Journal of Harbin Engineering University*, vol. 32, no. 9, pp. 1262–1266, 2011.

[20] S. Chen, *Some Modern Design Methods of Precise and Complex Structures*, Press of Beijing University of Aeronautics and Astronautics, Beijing, China, 1992.

[21] S. Chen, *Analysis, Synthesis and Optimization of Engineering Structural Systems*, China Science Culture Publishing House, Wan Chai, HongKong, 2008.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 303967, 10 pages
http://dx.doi.org/10.1155/2013/303967

Research Article

Formulation and Validation of Multidisciplinary Design Problem on Wear and Fatigue Life of Lead Screw Actuators

Krishna Meruva, Zhuming Bi, Donald Mueller, and Bongsu Kang

Department of Engineering, Indiana University-Purdue University Fort Wayne, Fort Wayne, IN 46805, USA

Correspondence should be addressed to Zhuming Bi; biz@ipfw.edu

Received 16 February 2013; Accepted 27 March 2013

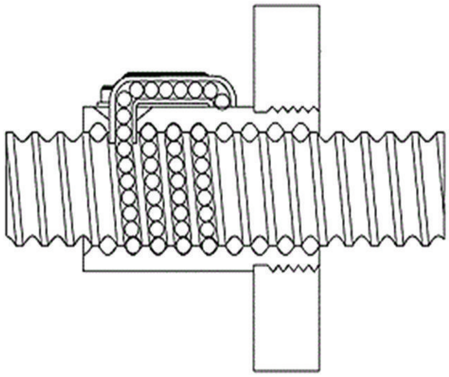
Academic Editor: Dan Zhang

Multidisciplinary design optimization has been widely applied in the optimization of large-scale complex system and also in the design and optimization of components, which are involved in multidisciplinary behaviors. The wear and fatigue life of lead screw actuators is a typical multidisciplinary problem. The wear behaviors of actuators closely relate to many factors such as loads, lubrications, materials properties, surface properties, pressures, and temperature. Therefore, the wear and fatigue life of actuators cannot be modeled without a simultaneous consideration of solid mechanics, fluid dynamics, contact mechanics, and thermal dynamics. In this paper, the wear and fatigue life of a lead screw actuator is modeled and validated. Firstly, the theory of asperity contact and Archard's model of sliding wear are applied to estimate the amount of wear under certain circumstances. Secondly, a test platform is developed based on a standard ASTM test protocol, and the wear phenomenon at the ball-on-flat sliding is measured to validate the developed wear model. Thirdly, finite element analysis is conducted using Nastran to assess the contact stresses in the lead screw and nut assembly model. The estimated data from the three sources are finally merged to formulate a mathematical model in predicting the wear and fatigue life for the optimization of lead screw actuators.

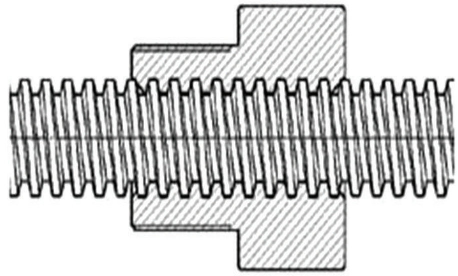
1. Introduction

Robotic technologies are playing more and more significant roles in our modern society. Robots are migrating from traditional manufacturing areas to our daily lives to assist or replace humans in completing tedious and routine tasks. Robots have gradually been applied in space, entertainment, military, agriculture, healthcare, and home environments. In answering the question about what emerging technology will cause another big stir in the way that the home computer did years ago, Bill Gates responded that robots will change the way people understand computers, and robots will interact with people in totally new ways [1]. However, many issues related in the field of robotics have not been addressed satisfactorily in terms of cost, precision, stiffness, reliability, energy efficiency, flexibility, robustness, and control. The ultimate goal of the presented work is to provide durable actuators for robotic design optimization from the perspective of cost and functionalities.

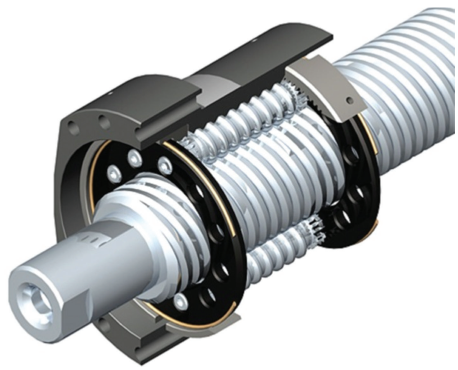
Robots are typical mechatronic products which need design knowledge from multiple disciplines such as mechanical engineering, electronic engineering, sensing technologies, artificial intelligence, and machine control [2]. Multidisciplinary design methodologies provide vital tools to support robotic design and applications. Since ro-botic systems are usually complex systems, decomposition and modularization is an effective way to manage system complexity [3–5]. Under a modularized architecture, robots can be treated as a set of basic modules, comprising rotary or linear modules as actuators and connecting modules for robotic structures. In particular, linear or rotary modules are essential to the majority of robots which need the capabilities of locomotion or manipulability. For examples, key components of a cutting-edge Exechon hybrid robot are three high-precision linear actuators [6, 7], while a medical robot with three low-cost linear actuators provides a 3 DOF motion for ankle rehabilitation [8].



(a) Ball screw (Roton, inc.)



(b) Lead (acme/trapezoidal) screw



(c) Roller screw (Thomson linear inc.)

Figure 1: Types of power screw mechanisms used on linear actuators.

In designing a robot, linear actuators are selected based on many aspects such as cost, precision, maximized loads, maximized speeds, and stiffness [9]. Most of the design criteria are in conflict with each other. However, it is our observation that the methodologies and data are lacking to help robotic designers select modular components in particular, when the design criteria of precision, cost, and fatigue life have to be considered simultaneously to optimize the system performance of a robot. The focus of this paper is design of the low-cost lead screw linear actuators. Multidisciplinary design problem will be formulated to explore the intricate relationships of loads, wear, and fatigue life. The rest of paper is organized as follows. In Section 2, the design of lead screw actuators will be overviewed. In Section 3, the adhesive friction models are developed to correspond the loads and stresses with the wear. In Section 4, the experiment is set up to acquire test data and to verify the developed models. In Section 5, the presented work is summarized.

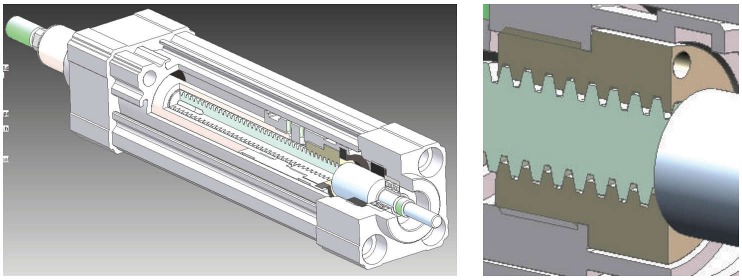
2. Overview of Lead Screw Actuators

A lead screw linear actuator is a mechanism that converts an input in the form of a rotary motion into a desired linear motion. The major benefits of using a lead screw mechanism in linear actuators are inherent mechanical advantage, high stiffness, high strength, and a cost-effective package. Lead screws fall under the category of power screws and can be classified into the following different types: *ball screw*, *acme/trapezoidal screw* and *roller screw*.

A ball screw, mechanism, as shown in Figure 1(a), consists of a ball screw and a ball nut with recirculating balls providing rolling contact between the nut and the screw. A large number of balls used in these assemblies facilitate the distribution of forces transmitted, resulting in a relatively small load per ball. These screws offer high efficiencies of around 90% due to the inherently low friction of the rolling elements.

An acme or trapezoidal screw, which hereafter will be addressed as lead screw, consists of a screw and a nut that are in sliding contact with each other. The screw is generally made up of alloy steel with a trapezoidal thread form, and the nut is typically made of an engineering polymer or bronze. The contact between the nut and the screw is a sliding contact; therefore friction plays a very important role in the performance and efficiency of the mechanism. These screws offer low efficiencies due to the relatively greater coefficient of friction in sliding.

As shown in Figure 1(c), planetary roller screws consist of a screw and a nut with planetary rollers. The barreled surfaces of all the rollers are used to transfer load from the nut to the screw [10]. Due to this increased load-bearing area, the roller screws can withstand higher dynamic loads with far greater life expectancy than either ball screw or lead screw mechanisms. In spite of the benefits that the ball and roller screws offer, they both tend to be far more expensive than



(a) Lead screw actuator

(b) detailed view of screw and nut

Figure 2: Lead screw cylinder (courtesy, PHD, Inc.).

lead screws. This is the reason why in applications where the operating speeds are lower and where life expectancy is not very high, lead screws offer a cost-effective design solution.

A typical single-axis linear actuator driven by a lead screw, as shown in Figure 2, consists of the following components: *a lead screw*, *lead nut*, *piston rod*, *cylinder tube*, *head*, *cap*, *radial ball bearing*, *electric motor*, *motor coupler*, *a mounting plate,* and other miscellaneous hardware such as fasteners.

2.1. Lead Screw Thread Geometry. Lead screws have either an Acme or Trapezoidal thread profile. The acme threads, which are used on inch size screws, have a 14.5° flank angle with a total included angle of 29° as shown in Figure 3. The trapezoidal screws, on the other hand, are used on metric size screws, having a 15° flank angle and an included angle of 30°.

Pitch and lead are closely related as shown in Figure 4. Pitch is the distance between the crests or troughs of two consecutive threads. Lead is the distance traversed along the axis of the screw, by one complete rotation, or 360°, of the screw:

$$\begin{aligned} &\text{single-start thread}, \ P_h = P \\ &\text{double-start}, \ P_h = 2P \\ &\text{three-start}, \ P_h = 3P. \end{aligned}$$

2.2. Screw Mechanics. The primary function of a lead screw is to traverse a load F axially, through a specified linear distance, L, called travel. It is a single-degree-of-freedom mechanism with the travel constrained between a full extent and full retract position. The nut, which is engaged with the screw, is generally the linearly traveling member, whereas the screw is the rotating member driven by a prime mover like an electric motor or by hand. Assuming that the screw is driven by an electric motor at an angular velocity of n, linear velocity v of the nut is given by

$$v = n \cdot P_h, \tag{1}$$

where v: m/s, n: revolutions/minute, and P_h: m.

A single-start threaded screw with pitch P and helix angle λ is shown in Figure 5. Consider that a single thread of the screw is unrolled for exactly one turn. The force diagram for lifting the load is shown in Figure 6(a), where the lifting force

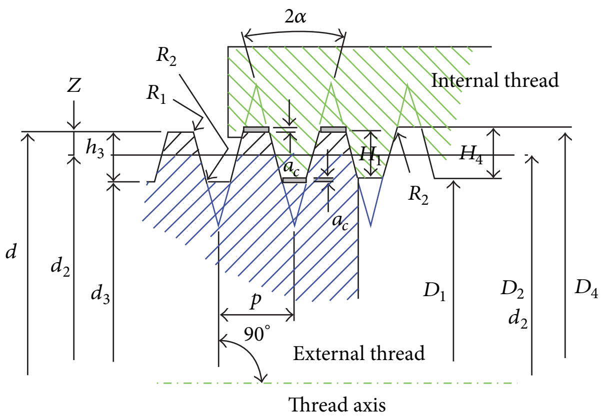


Figure 3: Trapezoidal thread profile, Din 103-4 (http://www.roymech.co.uk/). Legends: D_s: Nominal or major diameter of screw, D_n: nominal or major diameter of nut, K_s: minor diameter of screw, K_n: minor diameter of nut, E: pitch diameter, P: pitch, P_h: Lead, α: flank angle (°), h_s: depth of thread of screw, h_n: depth of thread of nut, h_e: depth of thread engagement.

F_r is positive. Similarly the force diagram for lowering the load is shown in Figure 6(b), where the lowering force F_l is negative [11]. The torque required to raise, T_r, or to lower, T_l, the load while overcoming the friction is given by

$$T_r = \frac{FE}{2}\left(\frac{\pi\mu E + P_h \cos\alpha_n}{\pi E\cos\alpha_n - \mu P_h}\right), \tag{2}$$

$$T_l = \frac{FE}{2}\left(\frac{\pi\mu E - P_h \cos\alpha_n}{\pi E\cos\alpha_n + \mu P_h}\right), \tag{3}$$

where λ = helix angle of the screw, μ = coefficient of friction (dry), α_n = flank angle measured in normal plane, $\alpha = 15°$ (metric trapezoidal thread), and n_s = number of starts;

$$\begin{aligned} P_h &= Pn_s, \\ \tan\alpha_n &= \tan\alpha\cos\lambda, \\ \tan\lambda &= \frac{P_h}{\pi E}. \end{aligned} \tag{4}$$

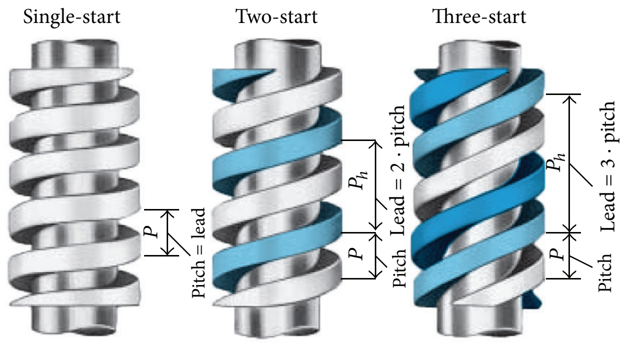


Figure 4: Singlestart and multistart threads (http://www.kammerer-gewinde.com/).

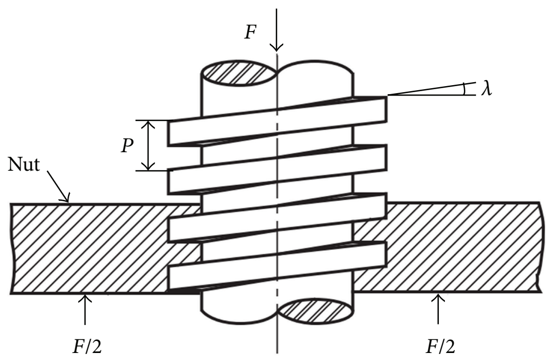


Figure 5: Free body diagram of lead screw [11].

One advantage of using a lead screw over other screws is its inherent ability to self-lock. This condition occurs when the torque required to lower the load T_l in (3) is greater than zero:

$$\pi \mu E > P_h \cos \alpha_n \quad \text{or} \quad \mu > \tan \lambda \cos \alpha_n. \tag{5}$$

The physical interpretation of (5) is that self-locking of a lead screw mechanism occurs when the coefficient of friction between the surfaces of the screw and nut in sliding contact is greater than the product $\tan \lambda \cos \alpha_n$. The screw efficiency is given by [12]:

$$\eta = \frac{\tan \lambda \ \left(1 - \mu \sec \alpha \tan \lambda\right)}{\left(\mu \sec \alpha + \tan \lambda\right)}. \tag{6}$$

From (6) it can be seen that the efficiency η of a lead screw mechanism is dependent on coefficient of friction between the two surfaces in contact, the helix angle λ and the flank angle. Out of these, the coefficient of friction is the most important factor influencing the performance of linear actuators using these mechanisms. Typical values of coefficient considered for different material combinations are given in (Table 1).

3. Adhesive Wear Model

Friction and the corresponding resultant wear between a metal and a nonmetal mating pair has been studied and researched a great deal by many in the past. However, due to the complexity of wear phenomenon, a unified, consistent, and accurate model does not exist for the material pair of metals and nonmetals. At the outset, it is important to understand the different types of wear, before attempting to formulate a model for a given case.

Table 1: Friction values for unlubricated common nut materials at room temperatures.

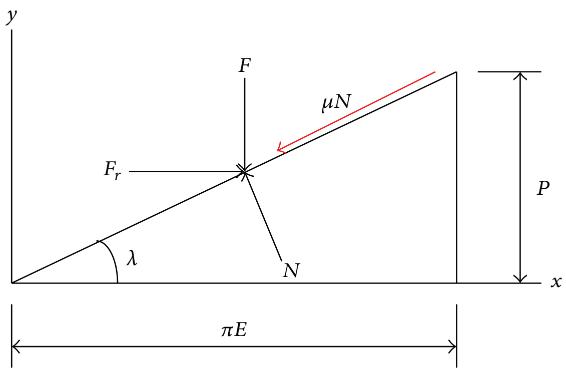
Screw	Nut	Coefficient of friction
AISI 1055 steel	Cast iron GC	0.18
	Steel	0.15
	Bronze CuSn	0.1
	Plastic	0.1

According to Burwell [13], there are four main forms of wear—adhesive, abrasive, corrosive, and surface fatigue wear. Adhesive wear is a type of wear between two soft surfaces in sliding contact with each other, where material is pulled off one surface and adheres to the other. An Abrasive wear occurs when a rough hard surface, or a soft surface with embedded hard particles on its surface, slides on a softer surface and causes a series of grooves on it. Corrosive wear is a process where wear occurs in a corrosive environment. Lastly Surface fatigue wear is observed during repeated sliding or rolling over a track. The repeated loading and unloading cycles to which the materials are exposed may induce the formation of surface or subsurface cracks, which eventually result in breakup of the surface with the formation of large pits on the surface [14].

Determining the type of wear responsible by examining a failed component is often a complex task [14]. Based on the studies conducted by Burwell [13]; Fuman et al. [15]; Love [16]; Ronan et al. [17], adhesive wear is believed to be the most common form of wear and also the least preventable. In all the mechanical systems where two solids are in sliding contact with each other, the adhesive wear is the primary form of wear. Although this form of wear is the most common, the wear rates are usually very low [14]. Only an unexpected occurrence of other forms of wear, for example, abrasive wear, often produces unexpectedly high wear rates and early failure of mechanism.

In particular adhesive wear occurs whenever one solid material is slid over the surface of another or is pressed against it. The wearing off of material takes the form of small particles or asperities, which are usually transferred to the other surface, but which may come off in loose form [14].

The process of adhesive wear can be explained with the help of Figure 7, wherein a softer part shown on top is sliding on a harder bottom part. Out of the two surfaces shown in the figure, the first surface is the physical interface between the two parts, and the second surface is a continuous surface inside the softer part on the top. When a tangential displacement is applied to one of the two parts parallel to the plane of the interface, shearing could happen along either of the two surfaces. If the force required to break through the interface of the materials is larger than the force required for breaking through some continuous surface



(a) (b)

Figure 6: (a) Force diagram for lifting the load and (b) for lowering the load.

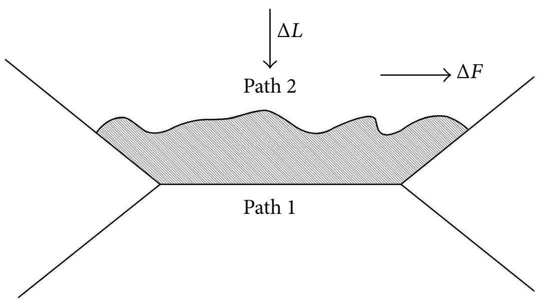


Figure 7: Adhesive junction being sheared.

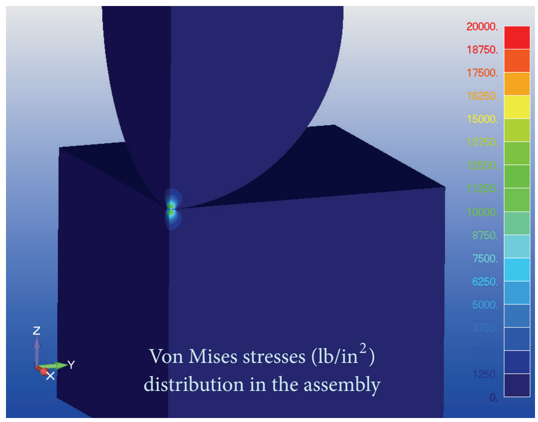


Figure 8: Ball-on-plate quarter FEA model.

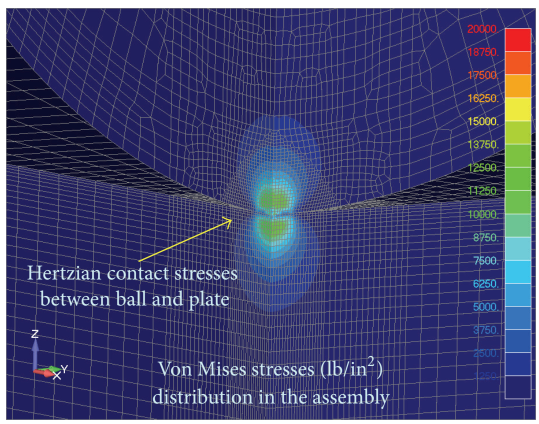


Figure 9: Contact stresses at the interface between ball and plate.

inside one of the materials, the break will occur along this latter surface. However, wear tests performed on different material combinations by Archard and Hirst showed that shearing along a surface inside of one the materials is very unlikely. Hence the break during shearing is most likely to occur at the interface.

The adhesive wear model, based on Archard's wear law, which is commonly used in practice, is utilized to compute theoretical wear:

$$W = K \cdot F \cdot V \cdot t, \tag{7}$$

where W = wear volume (in^3), K = wear factor (in^3-min/ft-lb-hr), F = nomal load (lb), V = sliding velocity (ft/min), t = elapsed time (hours).

Using this model, wear volume is calculated using a ball-on-flat sliding model. A 0.5 in. diameter ball made of 1060 steel is used to apply different normal loads on a flat plate made of Torlon PAI 4301 polymer. The contact stresses that act at the ball and plate junction are calculated, as listed in Table 2, using Hertzian contact stress model, which has been illustrated in Figure 16. These values are furthermore verified by performing finite element analysis of the assembly as shown in Figures 8 and 9. Also shown below in Figures 10 and 11 are the graphs representing the wear volume caused by sliding of the steel ball on plate.

4. Measurement and Validation of Wear

4.1. Sliding Wear Test. For determining the sliding wear of the polymer material used as the lead screw nut, a standard test setup using a linear, reciprocating ball-on-flat plane geometry is utilized [18]. The direction of the relative motion between

Table 2: Operating parameters used to compute wear volume.

Sample no.	Load F (lb)	Velocity V (ft/min)	Contact stress P (lb/in^2)	PV value (lb·ft/in^2·min)	Wear factor K (in^3·min/ft·lb·hr)
1	0.53	11.8	12038	142048	1.00E − 009
2	0.53		12038	142048	1.00E − 009
3	1.52		17114	201945	1.00E − 009
4	1.52		17114	201945	1.00E − 009
5	2.41		19870	234466	1.00E − 009
6	2.41		19870	234466	1.00E − 009

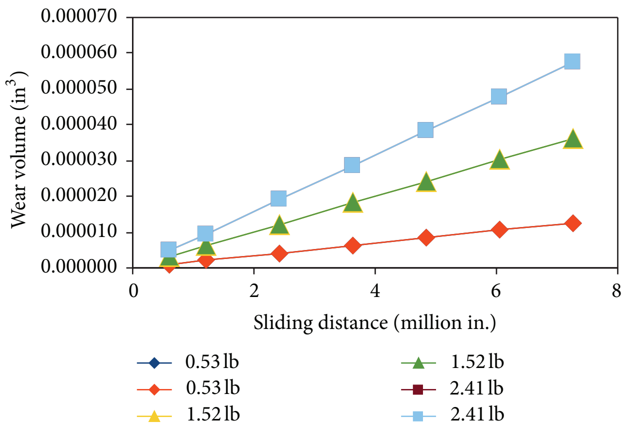


Figure 10: Graphs of wear against sliding distance for 1060 steel ball on Torlon 4301 polymer at various loads, velocity 11.8 ft/min.

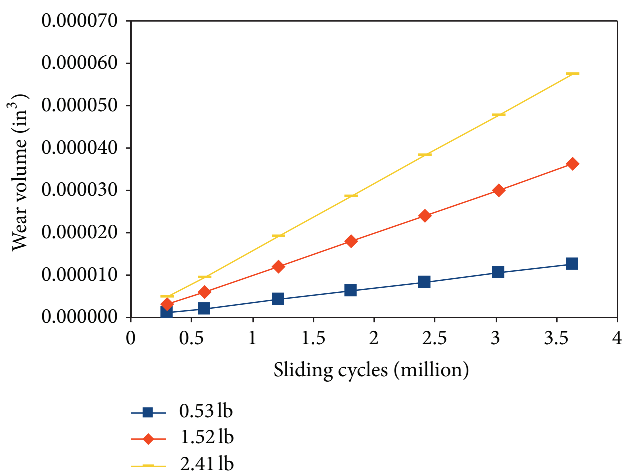


Figure 11: Graphs of wear against number of sliding cycles for 1060 steel ball on Torlon 4301 polymer at various loads, velocity 11.8 ft/min.

sliding surfaces reverses in a periodic fashion such that the sliding occurs back and forth in a straight line.

This test involves two specimens—a flat specimen made up of polymer material, namely, Torlon PAI, and a ball specimen made up of 1060 alloy steel, which slides against the flat specimen. The load is applied vertically downwards

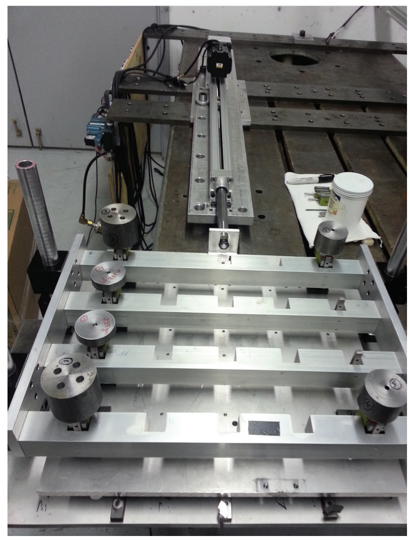


Figure 12: Wear test apparatus.

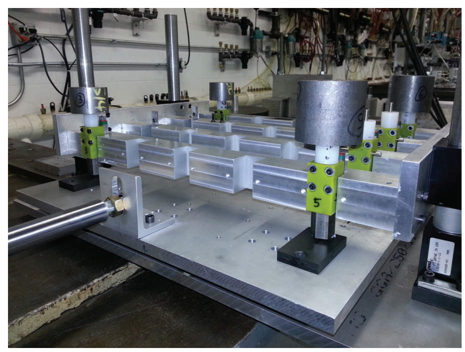


Figure 13: Ball and flat mounting arrangement.

through the ball specimen against the horizontally mounted flat specimen, as shown in Figures 12, 13, 14, and 15.

The loads selected in this test are 0.53 lb, 1.52 lb, and 2.41 lb. The average sliding velocity is set at 11.8 ft/min. The test apparatus as shown in Figure 12 consists of a fixture which contains a number of ball holders arranged at equal distances. A moving platform contains all the flat specimens securely fastened to it. The platform is provided with linearly reciprocating motion by a linear actuator powered by a servo motor.

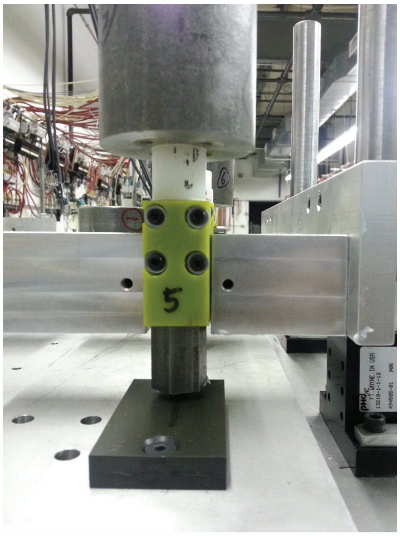


Figure 14: Wear on the flat specimen caused by the sliding ball.

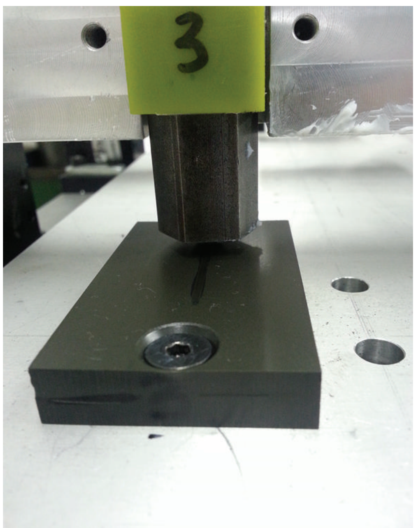


Figure 15: Closer view of the wear band on the flat.

The actuator is programmed to move the platform in a smooth reciprocating motion without any sudden acceleration and deceleration. The amount of travel in one direction is set to be 1 in., therefore a total travel of 2 inches per cycle. The rate at which the cycle is set to repeat itself is determined by the allowable duty cycle of the servo motor used in the test. The rate is therefore set at 30 cycles/minute.

Wear measurements are taken at periodic intervals—typically one to two weeks, using a dial indicator with a resolution of 0.0001 in. Additionally, wear profile is recorded using a profilometer. The measured values are shown in Table 4.

Comparison of the wear volumes between the wear model as shown in Table 3 and the measured wear values from the wear test shown in Table 4 exhibits a good correlation. The

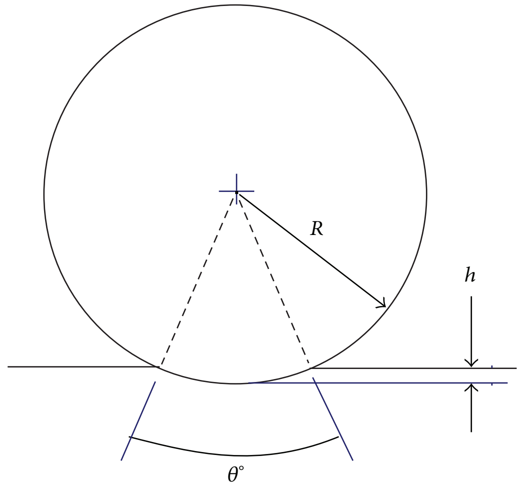


Figure 16: Ball and plate contact area.

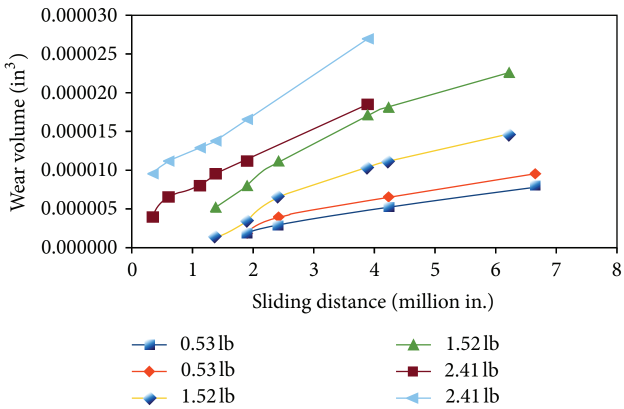


Figure 17: Graphs of measured values of wear against sliding distance for 1060 steel ball on Torlon 4301 polymer at various loads, velocity 11.8 ft/min.

values per the wear model are generally greater than the test values by a factor of 1.3 to 1.5.

4.2. Lead Screw Wear Model. The simulated wear test and the analytical wear model's results are thus used to generate a wear model for the lead screw actuator. In this process the first step is to evaluate the contact stresses that are developed at the lead screw and nut interface, by performing finite element analysis of the screw and nut assembly using NEi Nastran software. In order for the numerical solution in this analysis to converge, a very fine meshed model is required. But as a fine meshed model results in a large number of elements and the corresponding degrees of freedom, a symmetric quarter model is used, as shown in Figure 18. The assembly model used (as shown in Figure 18) as an example consists of a 12 mm diameter trapezoidal screw and nut assembly with a pitch of 3 mm. An axial load of 40 lb (177 N) is applied to the screw, whereas the nut is applied constraints to prevent translation in X-direction. Since a symmetric quarter model

Table 3: Wear volume as a function of sliding distance.

	Wear volume (in^3)						
Sliding distance							
Days	7	14	28	42	56	70	84
Hours	168	336	672	1008	1344	1680	2016
Million inches	0.6048	1.2096	2.4192	3.6288	4.8384	6.048	7.2576
Sample no.							
1	0.000001	0.000002	0.000004	0.000006	0.000008	0.000011	0.000013
2	0.000001	0.000002	0.000004	0.000006	0.000008	0.000011	0.000013
3	0.000003	0.000006	0.000012	0.000018	0.000024	0.000030	0.000036
4	0.000003	0.000006	0.000012	0.000018	0.000024	0.000030	0.000036
5	0.000005	0.000010	0.000019	0.000029	0.000038	0.000048	0.000057
6	0.000005	0.000010	0.000019	0.000029	0.000038	0.000048	0.000057

Table 4: Measured value of wear from the wear test for 1060 steel ball on Torlon 4301 polymer at various loads.

	Wear volume (in^3)							
Sliding distance								
(Million inches)	1.1232	1.3824	1.9008	2.4192	3.888	4.2336	6.2208	6.6528
Sample no.								
1			0.000002	0.000003		0.000005		0.000008
2			0.000002	0.000004		0.000007		0.000010
3		0.000001	0.000003	0.000007	0.000010	0.000011	0.000015	
4		0.000005	0.000008	0.000011	0.000017	0.000018	0.000023	
5	0.000008	0.000010	0.000011		0.000018			
6	0.000013	0.000015	0.000017		0.000027			

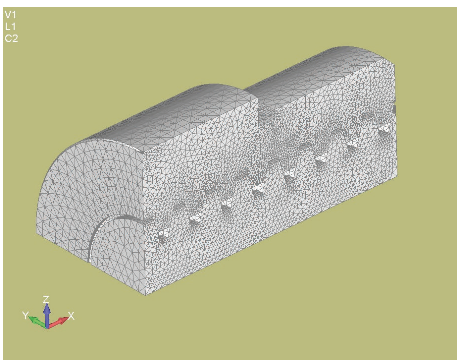


Figure 18: Symmetric quarter model of the lead screw and nut assembly.

of the assembly is analyzed, the load applied is one-fourth of the total load, which is 160 lb (708 N).

The results of the analysis are shown in graphical form in Figures 19 and 20. The distribution of von Mises stresses in the polymer nut and the mating steel screw are shown in Figures 19(a), 19(b), and 19(c). It can be seen in Figure 19(b) that the maximum stress in the nut is concentrated on the first thread, that bears the axial force transmitted by the screw, and also that this stress reduces in the subsequent threads in contact. An area of high stress concentration as seen in Figure 19(b) at the top left-hand corner of the nut should be ignored, since it is caused by the addition of a low-stiffness stabilizing spring element to stabilize the finite element model during analysis. Figures 20(a) and 20(b) show the total translation, caused by elastic deformation, in the assembly and the nut.

The second step in this process of generating a wear model for the lead screw actuator is to combine the wear formulation from Section 3 with the test results from Section 4 and apply a correction factor to the wear constant, K, and develop a wear equation of lead nut (8):

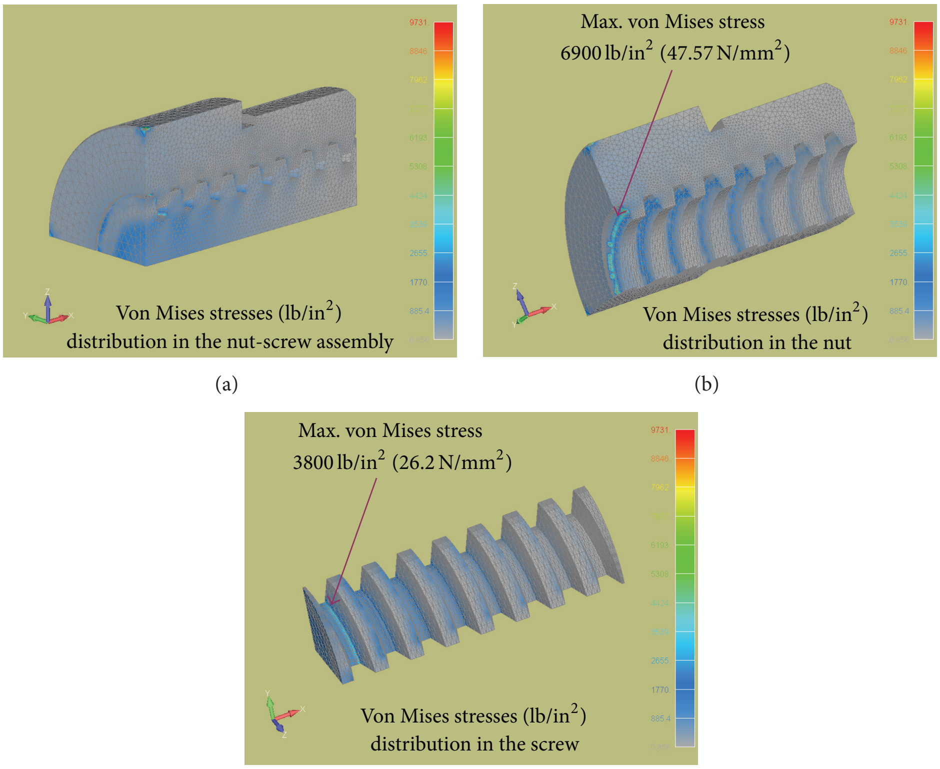
$$W = K_a F V t, \tag{8}$$

where K_a = adjusted wear contant (in^3-min/ft-lb-hr), F = normal load (lb), V = sliding velocity (ft/min), and t = total elapsed time (hr).

5. Conclusion

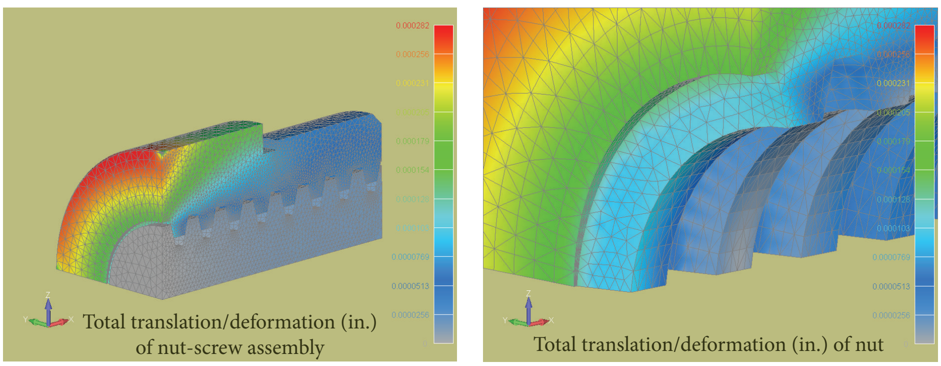
In this paper, the multidisciplinary design problem of lead screw linear actuators is discussed and formulated. The mathematical models of wear have been developed to correspond the load with fatigue life of actuators. The models have been validated by the test data and simulation data from finite element analysis. The presented work has its significance in predicting fatigue lives of low-cost linear actuators and contributing to the optimization of robotic design. The presented methodologies can be extended to deal with multidisciplinary designs of other machine components.

The adhesive sliding wear model based on Archard's wear law has proven to demonstrate very good results with



(a) (b) (c)

Figure 19: Von Mises stress distribution in (a) assembly, (b) nut, and (c) screw.



(a) (b)

Figure 20: Total translation (deformation) in the (a) assembly and (b) nut.

the material pair of steel and Torlon. The most important factor in this model is the wear factor, K, which is a function of the product of pressure and velocity of the sliding contact motion. For the material considered in this paper, Torlon PAI, the values of K, depending on the PV value, vary from 10×10^{-10} to 50×10^{-10} in^3-min/ft-lb-hr. The calculated volumetric wear, as shown in Table 3, is based on a K value of 10×10^{-10}. At this value, the estimated wear from tests and the calculated wear are very close to each other, as is evident from Figures 10 and 17. However, if the value of K considered is greater than 10×10^{-10}, then the theoretical and actual wear volume begins to drift apart. Therefore considering the right K value based on PV value is very important, for accurately predicting wear using this model.

Finally, in predicting wear and fatigue life for various loading conditions for lead screw actuators it is very important to consider the correct wear constant value, depending on the PV value of the actuator's application. Further studies could be conducted to eliminate the effect of PV value on the wear constant, and as proposed by Archard and Hirst [19], use a constant related to the probability per unit encounter of production of a wear particle.

Acknowledgments

The authors would like to gratefully acknowledge PHD, Inc., Fort Wayne, IN, USA, for supporting this work and providing the necessary resources required for modeling and verification of the wear model. The authors would also like to acknowledge and thank the effort of Mr. Puneeth Meruva, Fort Wayne, IN, USA, for his valuable assistance in conducting the research and help with the compilation and analysis of the data.

References

[1] Bill Gates, "Robots Will Be The Next World-Changing Technology," 2013, http://finance.yahoo.com/news/bill-gates-robots-next-world-134500596.html.

[2] Z. M. Bi and W. J. Zhang, "Concurrent optimal design of modular robotic configuration," *Journal of Robotic Systems*, vol. 18, no. 2, pp. 77–87, 2000.

[3] Z. M. Bi and W. J. Zhang, "Modularity technology in manufacturing: taxonomy and issues," *International Journal of Advanced Manufacturing Technology*, vol. 18, no. 5, pp. 381–390, 2001.

[4] Z. M. Bi, W. J. Zhang, I. M. Chen, and S. Y. T. Lang, "Automated geneartion of the D-H parameters for configuration design of modular manipulators," *Robotics and Computer-Integrated Manufacturing*, vol. 23, no. 5, pp. 553–562, 2007.

[5] Z. M. Bi, W. A. Gruver, W. J. Zhang, and S. Y. T. Lang, "Automated modeling of modular robotic configurations," *Robotics and Autonomous Systems*, vol. 54, no. 12, pp. 1015–1025, 2006.

[6] Z. M. Bi and Y. Jin, "Kinematic modeling of Exechon parallel kinematic machine," *Robotics and Computer-Integrated Manufacturing*, vol. 27, no. 1, pp. 186–193, 2011.

[7] Z. M. Bi and L. Wang, "Energy modeling of machine tool for optimization of machine setup," *IEEE Transaction on Automated Manufacturing Systems*, vol. 9, no. 3, pp. 607–613, 2012.

[8] Z. M. Bi, "Design of a spherical parallel kinematic machine for ankle rehabilitation," *Advanced Robotics*, vol. 27, no. 2, pp. 121–132, 2013.

[9] Z. M. Bi and B. Kang, "Enhancement of adaptability of parallel kinematic machines with an adjustable platform," *ASME Journal of Manufacturing Science and Engineering*, vol. 132, no. 6, Article ID 061016, 2010.

[10] SKF, "Roller screws," 2013, http://www.skf.com/files/779280.pdf.

[11] D. B. Marghitu, *Kinematic Chains and Machine Components Design*, Elsevier Academic Press Publication, 2005.

[12] V. B. Bhandari, *Design of Machine Elements*, Tata McGraw-Hill, 2008.

[13] J. T. Burwell, "Wear tests and service performance," in *Interpretation of Tests and Correlaion With Service*, pp. 88–140, American society for metals, Cleveland, Ohio, USA, 1957.

[14] E. Rabinowicz, *Friction and Wear of Materials*, Wily Interscience Publication, 1965.

[15] G. Fuman, F. Tobin, and K. H. Strauss, "Abrasives and wear," *Lubrication*, vol. 43, pp. 13–24, 1957.

[16] P. P. Love, "Diagnosis and analysis of plain bearing failures," *Wear*, vol. 1, no. 3, pp. 196–210, 1957.

[17] R. J. Ronan, G. R. Furman, and J. T. Bugbee, "Petroleum laboratory investigations," *Lubrication*, vol. 44, pp. 29–44, 1958.

[18] ASTM International, "Standard test method for linearly reciprocating ball-on-flat sliding wear," ASTM G 133-05, 2010.

[19] J. F. Archard and W. Hirst, "The wear of metals under unlubricated conditions," *Proceedings of the Royal Society A*, vol. 236, pp. 397–410, 1956.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 749256, 11 pages
http://dx.doi.org/10.1155/2013/749256

Research Article

Predatory Search Strategy Based on Swarm Intelligence for Continuous Optimization Problems

J. W. Wang,[1,2,3] H. F. Wang,[4] W. H. Ip,[5] K. Furuta,[3] T. Kanno,[3] and W. J. Zhang[2]

[1] *Complex Systems Research Center, East China University of Science and Technology, Shanghai 200237, China*
[2] *Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, Canada S7N 5A9*
[3] *Department of Systems Innovation, the University of Tokyo, Tokyo 113-8656, Japan*
[4] *Institute of Systems Engineering, Northeastern University, Shenyang 110114, China*
[5] *Department of Industrial and Systems Engineering, Hong Kong Polytechnic University, Kowloon, Hong Kong*

Correspondence should be addressed to W. J. Zhang; chris.zhang@usask.ca

Received 12 February 2013; Accepted 10 March 2013

Academic Editor: Zhuming Bi

We propose an approach to solve continuous variable optimization problems. The approach is based on the integration of predatory search strategy (PSS) and swarm intelligence technique. The integration is further based on two newly defined concepts proposed for the PSS, namely, "restriction" and "neighborhood," and takes the particle swarm optimization (PSO) algorithm as the local optimizer. The PSS is for the switch of exploitation and exploration (in particular by the adjustment of neighborhood), while the swarm intelligence technique is for searching the neighborhood. The proposed approach is thus named PSS-PSO. Five benchmarks are taken as test functions (including both unimodal and multimodal ones) to examine the effectiveness of the PSS-PSO with the seven well-known algorithms. The result of the test shows that the proposed approach PSS-PSO is superior to all the seven algorithms.

1. Introduction

As one of the key foundations of intelligent systems, computational intelligence is a strong tool to deal with challenging optimization problems. Most intelligent optimization methods are inspired from the biological system, such as Genetic algorithm [1], artificial immune systems [2], artificial neural network [3], ant colony optimization [4], particle swarm optimization [5, 6], culture algorithm [7], and colony location algorithm [8]. There are also methods learned from the artificial or man-made system, for example, simulated annealing [9] and fractal geometry [10]. Compared with the traditional optimization algorithms, the intelligent optimization algorithms have fewer restrictions to the objective function and are of computational efficiency and thus are widely used in industrial and social problems [11, 12].

However, to many real-world applications, the intelligent algorithms as mentioned earlier all encounter a common problem; that is, it is difficult to deal with the balance between exploitation and exploration. Exploration is the ability to test various regions in the problem space in order to locate a good optimum, hopefully the global one, and exploitation is the ability to concentrate on the search around a promising candidate solution in order to locate the optimum precisely [13]. Fast convergence velocity tends to result in the premature solution as opposed to the best solution.

Inspired by the predatory behavior of animals, Linhares [14, 15] proposed the predatory search strategy (PSS) to solve the previous problem; in particular, he used PSS to solve discrete variable optimization problems. This strategy is an individual-based searching strategy and has a good balance between exploitation and exploration.

On a general note, the balance of exploitation and exploration is not only a key problem for discrete variable optimization problems but also for the continuous variable optimization problems. A natural idea extending Linhares's work is to apply the PSS to the continuous variable optimization problem. However, the feature of individual-based searching does not allow the PSS to be straightforwardly applied to the continuous variable optimization problem. The basic motivation of this paper is to address this issue.

In this paper, we attempt to integrate the PSS with swarm intelligence techniques to solve continuous variable optimization problems. There are two major models in swarm intelligence, particle swarm optimization (PSO) and ant colony optimization (ACO). ACO is used for discrete problems, while the PSO is used for continuous problems. Therefore, in this paper, PSO is adopted to be incorporated with the PSS. The proposed approach may be called PSS-PSO. The PSS-PSO approach is tested by five benchmark examples of complex nonlinear functions. Compared with well-known PSO algorithms, the PSS-PSO is found to achieve a superior performance to the existing algorithms for continuous variable optimization problems.

In the following, we will first present the related works of PSS and PSO algorithms. Then, we will propose the basic idea of PSS-PSO and discuss the implementation of the PSS-PSO approach in Section 3, followed by the experiment and analysis in Section 4. Finally, there is a conclusion with discussion of future work.

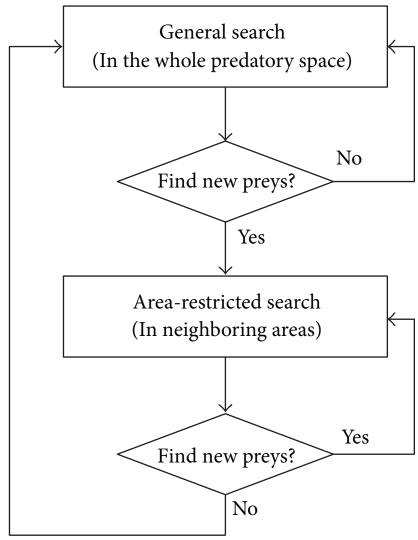


Figure 1: The predatory search strategy.

2. Related Studies

2.1. Predatory Search Strategy for Discrete Problems. Biologists have discovered that despite their various body constructions, the predatory search strategies of many search-intensive animals are amazingly similar [16–20]. When they seek food, they first search in a certain direction at a fast pace until they find a prey or sufficient evidence of it. Then, they slow down and intensify their search in the neighboring area in order to find more preys. After some time without success, they give up the intensified “area-restricted” search and go on to scan other areas [15]. As shown in Figure 1, this predatory search strategy of animals can be summarized by two search processes [21]: Search 1 (general search)—extensive search in the whole predatory space, and if prey or its proof is found, turn to search 2; Search 2 (area-restricted search)—intensive search in its neighboring area, and after a long time without success, turn to search 1.

This strategy is effective for many species because (besides being simple and general) it is able to strike a good balance between the exploitation (intensive search in a defined area) and the exploration (extensive search through many areas) of the search space. Linhares [14, 15] introduced the predatory search strategy as a new evolutionary computing technique to solve discrete optimization problems, such as TSP and VLSI layout. When this technique searches for the optimum, it first seeks solutions in the whole solution space until finding a “good” solution. Then, it intensifies the search in the vicinity of the “good” solution. After some time without further improvement, it gives up the area-restricted search and turns to the original extensive search. In this technique, a concept called *restriction* was defined as the solution cost to represent the neighborhood of a solution, which is used to adjust the neighboring area and implement the balance of exploitation and exploration.

The previous approach, in essence, can be viewed as a strategy to balance exploitation and exploration; however, it does not give any detail of the general search and area-restricted search. Hence, in this paper, we call it predatory search strategy (PSS), instead of predatory search algorithm. As mentioned in Section 1, the balance between local search and global search or exploration and exploitation is a key to all the intelligent algorithms. PSS is an example strategy to address this key. Liu and Wang [21] proposed another way to implement the PSS by adopting the solution distance as the restriction. All these studies focus on the discrete variable optimization problem.

On a general note, the general search enables the approach with a high search quality to avoid falling into the local optimum, while the area-restricted search enables the approach to have a fast convergence velocity. It is interesting to note that such search strategy was taken three decades ago in the M. S. degree thesis of Zhang [22]. In his approach, a random search technique and a gradient-based technique were combined to solve a specific continuous variable optimization problem for design of special mechanisms [22]. In his combined approach, the switch criterion for the two search algorithms was specifically related to the particular application problem; in particular, in the general search, the stop criterion is the feasible solutions in a highly constrained region [22]. The very motivation of the Zhang's combined approach was to overcome the difficulty in obtaining a feasible solution, because the constrained area was extremely “narrow.” Zhang's work was at times prior to the development of intelligent optimization ideas, concepts, and techniques such as PSO. His work is inspiring but not a generalized one.

2.2. Particle Swarm Optimization and Its Developments. The PSO algorithm was first developed by Kennedy and Eberhart based on the simulation of a simplified social model [5, 6, 23]. The standard PSO algorithm can be explained as follows.

A swarm being made up by m particles searches a D-dimensional problem space. Each particle is assigned a randomized velocity and a stochastic position. The position represents the solution of the problem. When “flying” each particle is attracted by a good location achieved so far by itself and by a good location achieved by the members in the whole swarm (or the members in the neighborhood). The position of the ith particle is represented as $x_i = (x_{i1}, \ldots, x_{id} \ldots, x_{iD})$, and its velocity is represented as $v_i = (v_{i1}, \ldots, v_{id}, \ldots, v_{iD})$, $1 \le i \le m$, $1 \le d \le D$. The best previous position of the ith particle, namely, the position with the best fitness value, is represented as $p_i = (p_{i1}, \ldots, p_{id}, \ldots, p_{iD})$, and the index of the best particle among all the particles in the neighborhood is represented by the symbol g. Each particle updates its velocity and position according to the following equations:

$$v_{id}^{k+1} = \omega v_{id}^{k} + c_1 \xi \left(p_{id}^{k} - x_{id}^{k}\right) + c_2 \eta \left(p_{gd}^{k} - x_{id}^{k}\right), \qquad x_{id}^{k+1} = x_{id}^{k} + v_{id}^{k+1}, \tag{1}$$

where ω is the inertia weight that determines how much a particle holds its current velocity in the next iteration. c_1 and c_2 are learning factors, also named acceleration constants, which are two positive constants. Learning factors are usually equal to 2, while other settings can also be seen in the literature [13]. ξ and η are pseudorandom numbers and they obey the same homogeneous distribution in the range $[0, 1]$. The velocity of a particle is limited in the range of $V_{\max}$. $V_{\max}$ is set to be the range of each dimension variable and used to initialize the velocity of particles without selecting and tuning in detail. The running of the PSO is much similar to evolutionary algorithms such as GA, including initialization, fitness evaluation, update of velocity and position, and testing of the stop criterion. When the neighborhood of a particle is the whole population, it is called the global version of PSO (GPSO) [24]; otherwise, it is called local version of PSO (LPSO) [25].

The PSO algorithm has increasingly attracted attention now [13, 24, 26, 27] with many successful applications in real-world optimization problems [28–33]. The most salient advantage of PSO is its fast convergence to the optimum; however, this feature also tends to lead the whole swarm into the local optimum, especially for solving multimodal problems [34].

Different strategies have been developed in the literature to solve the problem with PSO. Liang et al. [34] proposed a comprehensive-learning PSO (CLPSO) which uses a novel learning strategy where the historical best information of all other particles is used to update a particle's velocity for multimodal applications. Shubham et al. [35] proposed a fuzzy clustering-based particle swarm (FCPSO) algorithm by using an external repository to preserve nondominated particles found along the search process to solve multiobjective optimization problems. Zhan et al. [36] proposed an adaptive PSO (APSO) by developing a systematic parameter adaptation scheme and an elitist learning strategy. Clerc [37] proposed an adaptive PSO, called TRIBES, in which the parameters change according to the swarm behavior. TRIBES is a totally parameter free algorithm and the users only modify the adaptive rules [37, 38]. In the work of [39], the flight mechanism of wild goose team, including goose role division, parallel principle, aggregate principle, and separate principle, was proposed to address the problem with PSO [39]. Silva et al. [40] proposed a variant of PSO, called predatory prey optimizer (PPO), by introducing another swarm; in particular, original swarm is called prey swarm and the new swarm is called predator swarm. These two swarms have different dynamic behaviors which will balance the exploration and exploitation in PSO. They further proposed a simple adaptable PPO by introducing a symbiosis adaptive scheme into PPO [41]. The difference between their work and ours will be discussed in the next section after some detail of our approach is introduced.

3. The Proposed PSS-PSO Approach

3.1. Basic Idea. To make the PSS be capable of solving continuous variable optimization problems, we need to address three issues: (1) how to search the continuous space, including the general search and the area-restricted search, (2) how to define restrictions, and (3) how to define the neighborhood. The first issue can be addressed by the PSO; so we will not discuss it here. In the following, we will address the last two issues. We start with proposing two concepts, namely *Restriction* and *Neighborhood*.

Restriction. It is the distance between two points in a multidimensional space. Here, we use the Euclid norm to define the distance.

Neighborhood. The neighborhood of a point, or a particle, under some restriction, is defined as a hyperspherical space which takes the particle as the center and the restriction as the radius.

Further, there are Lrestrictions in the initial solution space, which are represented as $restriction(0), \ldots, restriction(l), \ldots, restriction(L-1)$, where $l \in \{0, 1, \ldots, L\}$ is called the level of the restriction. Therefore, the neighborhood of a point, say x, under the restriction, $restriction(l)$, can be represented as $N(x, restriction(l))$.

The list of the restriction levels needs to be recomputed when a new best overall solution, b, is found, because a restricted search area should be defined around this new improving point. To build such a restriction list, the following scheme is used: initialize L particles in the initialization space and compute each distance between one particle and b. Rank the Ldistances. It is noted that $restriction(0)$ carries the smallest distance, and $restriction(1)$ carries the second smallest distance, and so on.

The overall procedure of PSS is given as follows.

Choose one random point x in the initial space ψ. A swarm consisting of m particles is initialized in $N(x, restriction(0))$, and then the search is carried out with the standard PSO update equations. The number of iterations is controlled by variable n. If no better solution can be found, the swarm will be initialized in $N(x, restriction(1))$, and repeat the search; if a better solution, b, is found, then the best solution will be updated by this new solution, and all the restrictions will be set up again. The neighborhood is enlarged or reduced through the adjustment of restrictions, which makes sense for the concept of balancing exploration and exploitation.

```
Begin
Choose x randomly from ψ
while l < L
        initialize m particles in N (x, restriction(l))
        for (i = 0; i < I; i + +)
              for (c = 0; c < n; c + +)
                    update swarm based on standard PSO equations
                    if g < p then p = g
              end for
              x = p
              if f(x) < f(b) then
              b = x, l = 0, i = 0, re-compute L restrictions
        end for
        l = l + 1
        if l = [L/5] then
              l = L − [L/S]
End while
```

Algorithm 1: Pseudocode of the PSS-PSO algorithm.

As mentioned in Section 2.2, Silva et al. [40, 41] proposed algorithms with similar names of ours. However, our idea is totally different from theirs. In their work, they used two swarms in PSO and each swarm has different behavior. The particles in the prey swarm have the same behavior with the standard PSO; the particles in the predatory swarm are attracted to the best individuals in the swarm, while the other particles are repelled by their presence [40, 41]. The essence of such a mechanism with their approach is a multipopulation strategy to keep the diversity of the whole swarm. Our idea only considers the behavior of the predator and the balance of exploration and exploitation is achieved by the movement of predatory individuals not a predatory swarm. Furthermore, there is only one swarm performing the local search in our algorithm PSS-PSO, while there are two swarms in their algorithm PPO.

3.2. Implementation. The pseudocode of the PSS-PSO algorithm is presented in Algorithm 1.

Let

ψ: The initial space;

L: The number of restrictions;

l: The level of the restriction, $l \in \{0, 1, \ldots, L\}$;

n: The maximum number of iterations of PSO;

c: The iteration number of PSO;

I: The maximum running times of PSO;

i: The running number of PSO;

g: Best solution of PSO in each iteration;

m: Number of particles in a swarm;

$N(x, restriction(l))$: The neighborhood of x, under the restriction restriction(l), $l \in \{0, 1, \ldots, L\}$;

b: A new best overall solution in ψ;

p: The best solution in $N(x, restriction(l))$.

It is noted that we used the integer division of L as the number of restriction levels for each search mode. The algorithm was made to stop when $[L/5]$ levels in general search are tried without finding a new better solution.

4. Experiments

4.1. Experimental Setting. Five benchmarks [13, 24, 42] which are commonly used in the evolutionary computation are selected as examples in this paper to test the proposed approach. The detail of the functions, including the formulations, the numbers of dimensions, and the search range of the variables, are summarized in Table 1. It is noted that in the experiment, the initial range of algorithm is set to be the same with the search range. The Sphere function and the Rosenbrock function are two unimodal functions and have only one global minimum in the initial search space. The Rosenbrock function changes very slightly in the long and narrow area close to the global minimum. The Schaffer's f_6 function, the Rastrigrin functions, and the Griewank function are three multimodal functions, which have very complicated landforms. The global optimums of the five functions are all zero. The five functions are quite fit for the evaluation of algorithm performance [42].

Seven existing PSO algorithms, as listed in Table 2, are compared with PSS-PSO. First three PSO algorithms, including GPSO, LPSO with ring neighborhood, and VPSO with von Neumann neighborhood, are well known and have widely been used in computation comparisons and real-world applications. The other three PSO algorithms are recent versions as mentioned in Section 2.2. CLPSO adopts a novel learning strategy and aims at better performance for multimodal functions. GTO refers to the goose team flight mechanism and the gradient information of functions and obtains fast convergence for the unimodal functions and some multimodal functions. APSO offers a systematic parameter adaptation scheme which can speed up other PSO variants. TRIBES

Table 1: Benchmarks.

Function	Formula	Dim	Search range
Sphere	$f_1(\vec{x}) = \sum_{i=1}^{n} x_i^2$	30	$[-100, 100]^n$
Rosenbrock	$f_2(\vec{x}) = \sum_{i=1}^{n-1} \left(100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2\right)$	30	$[-30, 30]^n$
Schaffer's f_6	$f_3(\vec{x}) = 0.5 + \dfrac{\left(\sin\sqrt{x_1^2 + x_2^2}\right)^2 - 0.5}{\left(1 + 0.001\left(x_1^2 + x_2^2\right)\right)^2}$	2	$[-100, 100]^2$
Rastrigrin	$f_4(\vec{x}) = \sum_{i=1}^{n} \left(x_i^2 - 10\cos(2\pi x_i) + 10\right)$	30	$[-5.12, 5.12]^n$
Griewank	$f_5(\vec{x}) = \dfrac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n}\cos\left(\dfrac{x_i}{\sqrt{i}}\right) + 1$	30	$[-600, 600]^n$

Table 2: PSO algorithms used in the comparison.

Algorithm	Reference
GPSO	Shi and Eberhart, 1998 [24]
LPSO	Kennedy and Mendes, 2002 [25]
VPSO	Kennedy and Mendes, 2002 [25]
CLPSO	Liang et al., 2006 [34]
GTO	J. Wang and D. Wang, 2008 [39]
APSO	Zhan et al., 2009 [36]
TRIBES	Cooren et al., 2009 [38]

is an adaptive algorithm which enables the users to be free in the parameter selection.

First, we will compare the PSS-PSO with the first six PSO algorithms on the ability of exploitation and exploration with the five benchmarks listed in Table 1. Then, we will compare the PSS-PSO with TRIBES on the four shifted benchmarks. GPSO, GTO, and PSS are programmed by Java run on a PC with P4-2.66 GHz CPU and 512 MB memory. Results of the other algorithms are from the literature, respectively. To get rid of the randomicity, the results are the average of 30 trial runs [42]. To test exploitation and exploration, we use different swarm sizes and stop criterions, which are given in the following discussion.

4.2. Comparisons on the Exploitation

4.2.1. Comparisons with Different Algorithms. For a fair comparison, the PSS and the first six PSO algorithms are tested with the same population size (m) of 20. The stop criterion is set according to reference [42] as number of function evaluations (FEs) = 200000. With consideration of the population size, this stop criterion also means that the number of iteration is 10000. In the following discussion, we will use the iteration number as the stop criterion. Here, we define the global optimum zero as "less than 10^{-45}." If the algorithm reaches the global optimum, it also ends. It is noted that, similar to the PSS for discrete problem, the PSS-PSO for continuous problems uses the restriction level as the stop criterion (see Algorithm 1). To compare with other PSO algorithms, we also record the iteration number and stop PSS-PSO when the iteration number reaches 10000.

The average optimums obtained by the seven algorithms are listed in Table 3. The data of PSS and GTO, are obtained by our experiments. Other data are from the references listed in Table 2. Symbol "—" in Table 3 means the data unavailable. It is noted that the results of CLPSO and APSO on the Rosenbrock function are with symbol "∗". The reason is that the two algorithms are not good at dealing with unimodal functions, particularly with the Rosenbrock functions. The initial ranges of the Rosenbrock functions for CLPSO and APSO are listed in Table 4 [34, 36], and therefore, the common idea is to narrow the search space of this function.

On the simple unimodal function Sphere, PSS-PSO, GPSO, GTO and APSO can all find the global optimum. Actually, GTO is the fastest one, which will be shown from the comparison on the exploration in the next section. On the complicated unimodal function Rosenbrock, PSS-PSO is the best one. Although GTO references the gradient information in the optimization process, the complicated search space still prevents it from obtaining the satisfactory result. On the simple multimodal function Schaffer's f_6, PSS-PSO, GPSO, and GTO can all reach the global optimum. On the complicated multimodal functions, Rastrigrin and Griewank, PSS-PSO, CLSPO and APSO are all good optimizers. PSS-PSO and CLPSO have better performance on Griewank; CLSPO and APSO have better performance on Rastrigrin. The results prove the theorem of "no free lunch" that no algorithm can outperform all the others on every aspects or every kind of problems [43]. However, the PSS-PSO outperforms the six PSO algorithms on most of the benchmark functions except the Rastrigrin function.

4.2.2. Analysis of the Search Behaviors. To further observe the search behaviors of the PSS-PSO, we try other two population sizes, 10 and 30, respectively, and extend the maximum iteration numbers to 50000 in the experiments and give the average convergence curves of PSS-PSO and GPSO on the five functions in Figure 2. In this figure, the x coordinate is the iteration number. The y coordinate is $\lg f(x)$, the common logarithm of the fitness value, because the fitness value changes too much in the optimization process.

(1) Exploitation for Unimodal Functions. From the convergence curves on the Sphere function, we can see that GPSO has a faster convergence speed. The reason is that Sphere

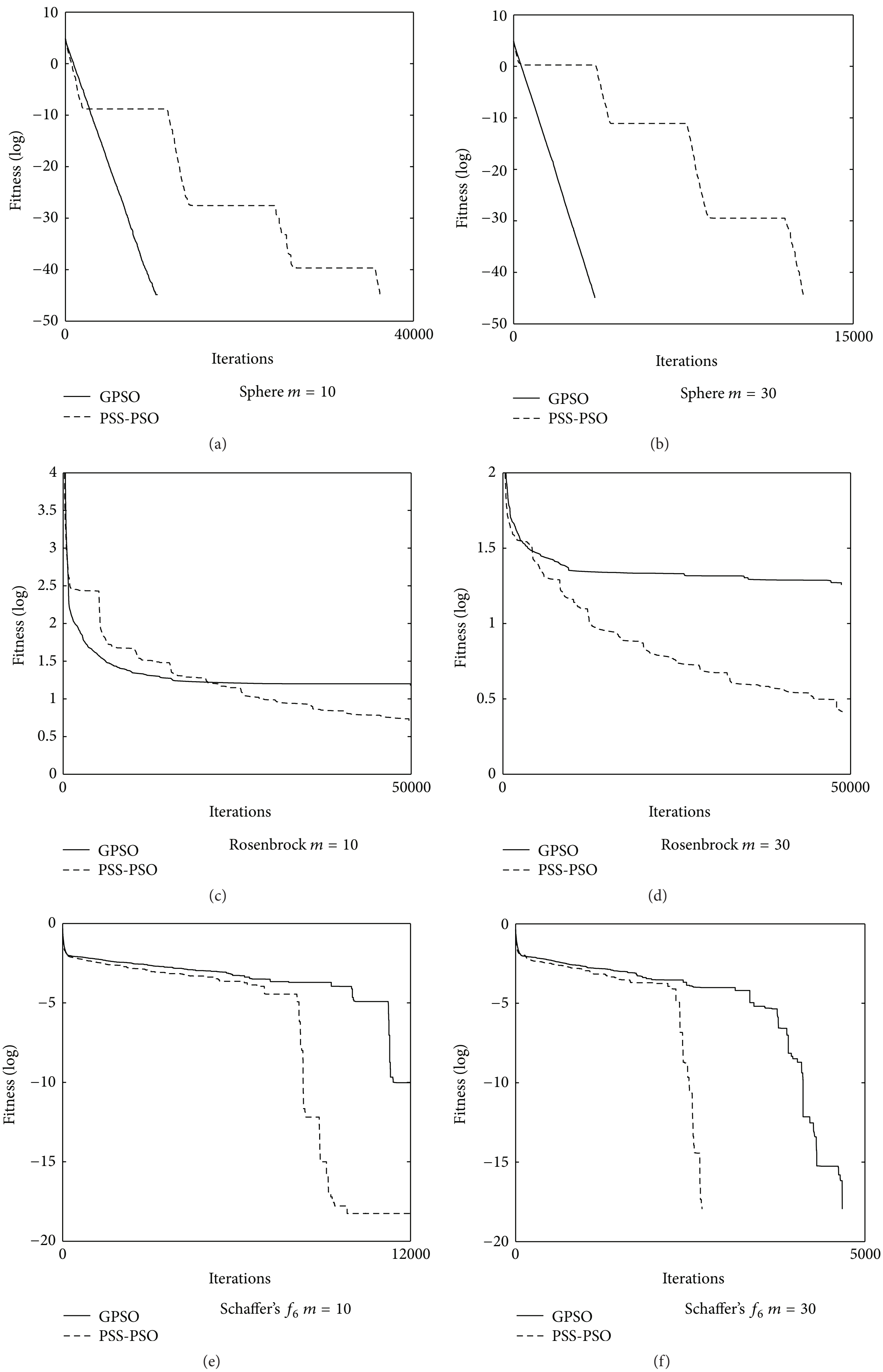


Figure 2: Continued.

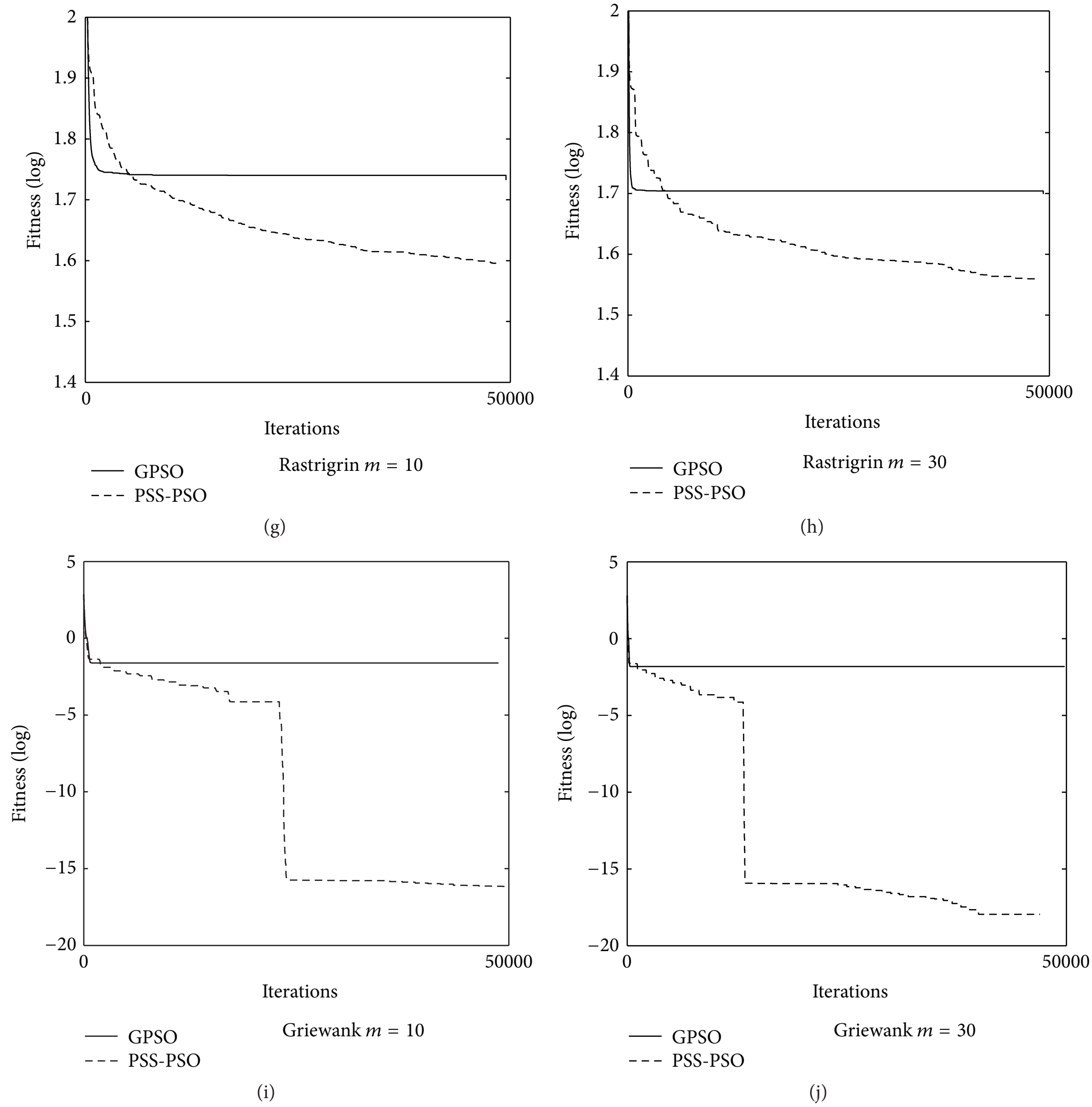


Figure 2: Average convergence curves of GPSO and PSS-PSO over 30 trials.

Table 3: Average optimums of seven algorithms.

Algorithm	Sphere	Rosenbrock	Schaffer's f_6	Rastrigrin	Griewank
PSS-PSO	0	0.0813065	0	13.2057	$1.0214E-16$
GPSO	0	24.486	0	29.6099	$2.37E-2$
LPSO	$6.33E-29$	18.9472	—	35.0819	$1.1E-2$
VPSO	$8.49E-39$	35.9005	—	27.7394	$1.31E-2$
CLPSO	$1.06E019$	11^{*}	—	$7.4E-11$	$6.45E-13$
GTO	0	5.1711316	0	215.24	$5.7518E-4$
APSO	0	2.84^{*}	—	$5.8E-15$	$1.67E-2$

Table 4: Search ranges of Rosenbrock for CLPSO and APSO.

Algorithm	Search range
CLPSO	$[-2.048, 2.048]^n$
APSO	$[-10, 10]^n$

Table 5: Goal for exploration.

Function	Sphere	Rosenbrock	Schaffer's f_6	Rastrigrin	Griewank
Goal	0.01	100	10^{-5}	100	0.1

is a simple unimodal function and the particles easily find the global optimum. On the optimization of the Rosenbrock functions, the problem of GPSO appears (i.e., not having a sustaining search ability during long iterations); in particular, GPSO cannot find better solutions after about 10000 iterations. On the contrary, PSS-PSO can obtain better solutions during the whole optimization process. The reason is that, as mentioned in Section 4.1, the Rosenbrock function changes very slightly in the long and narrow area close to the global minimum. This feature makes GPSO lose its ability of dynamic adjustment of particle's velocity, which leads the whole swarm to lose the energy of reaching better solutions. However, the switch of different restriction levels enables the swarm in the PSS-PSO with its ability of dynamic adjustment of particle's velocity continuously.

(2) Exploitation for Multimodal Functions. On the simple multimodal function Schaffer's f_6, GPSO and PSS-PSO can both perform well. However, on the two complicated functions, Rastrigrin and Griewank, GPSO exposed its deficiency again. After very early iterations, the curve of GPSO stays horizontal, which means that the whole swarm has been trapped into a local optimum. However, PSS-PSO does not have this problem, and the curves of PSS-PSO go down continuously.

4.3. Comparisons on the Exploration. To test the exploration of PSS-PSO, we use the following criterion in the experiment: whether the algorithm can find the goal set for each function in 4000 iterations. Using this criterion, we can get success rate, average iteration number, and average running time. Here, average iteration number is the iteration number for the successful runs; success rate is defined as number of successful runs/total number of runs. The predefined accuracy levels, namely, the goals of the different functions, are listed in Table 5. Since exploration is the ability to test various regions in the problem space in order to locate a good optimum, what we care about in this experiments is the convergence speed. Therefore, the goal value here is different from the global optimum, and in particular, the goal is an acceptable good solution. We compare PSS-PSO with GPSO and GTO in the experiments. We try two population sizes, 10 and 30, respectively, and run the algorithms 100 times for each case. The results are listed in Table 6.

The results show that even with a very small population size 10, PSS can still achieve 100% success rates for all the functions. As the standard version of PSO, GPSO has the worst performance, which is not beyond our expectation. On the three functions, Sphere, Schaffer's f_6, and Griewank, GTO can reach the goals very fast in even 1 or 3 iterations. This fast convergence speed of GTO benefits from the gradient information. However, on the two complicated functions, Rosenbrock and Rastrigrin, GTO does not have such good performance. It even fails to reach the goal of the Rastrigrin function at any trial. The comparison shows that PSS-PSO has an overall good performance on the exploration.

4.4. Comparisons on the Shifted Benchmarks. In the previous experiments, all the benchmarks are traditional functions with the global optimum at the origin of the search space, which may be more easily for some algorithms to find the global optimum [44]. Therefore, we use the shifted versions of benchmarks in Table 1 to perform a further test; in particular, the results are compared with that of TRIBES from the literature [38]. The shifted benchmarks are defined by Suganthan et al. [42] and labeled as F1 (shifted Sphere), F6 (shifted Rosenbrock), F7 (shifted rotated Griewank without bound), F10 (shifted rotated Rastrigrin), and F14 (shifted rotated expanded Schaffer's f_6). Because TIBES can not give satisfying result of F14 [38], we did not consider F14 in this experiment. Two tests are performed to examine the exploitation and exploration of two algorithms. To examine the exploitation, the algorithm stops if the number of evaluations of the objective functions exceeds $10000 * D$. The dimension D is set as 10. Each algorithm runs 25 times for one case. Then, we recorded the mean error for each benchmark. To test the exploration, we examine the number of function evaluations the algorithms needs to achieve a goal. The algorithm stops if it reachs the goal or if the number of evaluations of the objective functions exceeds $10000 * D$. The dimension D is also set as 10. Each algorithm runs 25 times for one case. The goals of functions, the mean of the number of functions evaluations (mean iterations for short), the success rates, and the performance rates are recorded. The performance rate is defined as mean iterations/success rate. All the results are listed in Table 7.

From the results in Table 7, we can see that the PSS-PSO has similar performance with TRIBES in F1, F6, and F7. PSS-PSO and TRIBES can both find global optimum in F1, which is a unimodel function, and PSS-PSO is a little slower than TRIBES. PSS-PSO and TRIBES both have very low ability to reach the goals for F6 and F7. In the runs that find the goals, PSS-PSO needs fewer number of iterations.

However, PSS-PSO has a much better result than TRIBES in F10. TRIBES cannot reach the goal of F10; on contrary, PSS-PSO has a success rate of 0.12. Furthermore, PSS-PSO can achieve a much less mean error in F10.

5. Conclusions

In this paper, we proposed an approach to integrate PSS with PSO, and the approach was named PSS-PSO. This integration is achieved by proposing two concepts with the PSS, namely restriction and neighborhood. Five benchmarks are taken as test functions to examine the exploration and

Table 6: Exploration performance of the different algorithms.

m	Algorithm	Index	Sphere	Rosenbrock	Schaffer's f_6	Rastrigrin	Griewank
10	GPSO	Suc. rate	1	0.82	0.75	0.97	0.96
		Iterations	1379.95	1847.4512	1808.3334	605.8866	1149.5938
		Ave. time (s)	0.216	0.360	0.059	0.154	0.264
	GTO	Suc. rate	1	0.98	1	0	1
		Iterations	1.04	164.79	3.2333333	—	1.01
		Ave. time (s)	0.009	0.038	0.011	—	0.009
	PSS-PSO	Suc. rate	1	1	1	1	1
		Iterations	832.66	1879.39	1570.9	690.5	998.34
		Ave. time (s)	0.140	0.317	0.040	0.144	0.200
30	GPSO	Suc. rate	1	0.93	1	1	0.99
		Iterations	320.59	780.17896	959.46	226.89	433.21213
		Ave. time (s)	0.159	0.464	0.058	0.144	0.292
	GTO	Suc. rate	1	1	1	0	1
		Iterations	1.06	111.31	3.2666	—	1
		Ave. time (s)	0.01	0.038	0.011	—	0.01
	PSS-PSO	Suc. rate	1	1	1	1	1
		Iterations	332.81	648.07	434.0	243.94	319.57
		Ave. time (s)	0.163	0.311	0.030	0.151	0.192

Table 7: Performance on shifted benchmarks.

Index	Algorithm	F1	F6	F7	F10
Accuracy		$1e-06$	$1e-02$	$1e-02$	$1e-02$
Mean iterations	PSS-PSO	4366	77290	62600	61600
	TRIBES	1364.72	98309.28	96568.44	$1E+05$
Success. rate	PSS-PSO	1	0.04	0.04	0.12
	TRIBES	1	0..04	0.04	0
Performance rate.	PSS-PSO	4366	1932250	1565000	513333.33
	TRIBES	1364.74	2457725	2414211	—
Mean error	PSS-PSO	0	9.692547	0.05717346	1.8716955
	TRIBES	0	0.85882	0.077474	12.118002

exploitation of the proposed approach. Experiments show that the ability of balancing exploration and exploitation makes the proposed approach be applicable for both the unimodal and multimodal functions. Compared with six existing PSO algorithms, PSS-PSO has achieved a better performance overall.

Some future research needs to be carried out on the proposed approach. First, we will test the PSS-PSO for more benchmarks, especially for more rotated functions. Second, we will test the PSS-PSO in complex environments; in particular, we will use it to solve the dynamic problems and multiple objective optimization problems. Third, we will use the PSS-PSO to solve the complex real-world applications such as the mixed topology and geometry optimization design problems in robotics [45].

Acknowledgments

The authors want to thank the financial support to this work by NSERC through a strategic project grant, and a grant of ECUST through a program of the Fundamental Research Funds for the Central Universities to W. J. Zhang, and by NSFC (Grant No. 71001018, 61273031) of China to H. F Wang. Our gratitude is also extended to the Department of Industrial and Systems Engineering of the Hong Kong Polytechnic University (G-YK04).

References

[1] J. H. Holland, *Adaptation in Nature and Artificial Systems*, University of Michigan Press, Ann Arbor, Mich, USA, 1975.

[2] J. D. Farmer, N. H. Packard, and A. S. Perelson, "The immune system, adaptation, and machine learning," *Physica D*, vol. 22, no. 1–3, pp. 187–204, 1986.

[3] J. J. Hopfield and D. W. Tank, "Neural computation of decisons in optimization problems," *Biological Cybernetics*, vol. 52, no. 3, pp. 141–152, 1985.

[4] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 26, no. 1, pp. 29–41, 1996.

[5] R. Eberhart and J. Kennedy, "New optimizer using particle swarm theory," in *Proceedings of the 6th International Symposium on Micro Machine and Human Science*, pp. 39–43, Indianapolis, Ind, USA, October 1995.

[6] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, Perth, Australia, December 1995.

[7] R. G. Reynolds and S. R. Rolnick, "Learning the parameters for a gradient-based approach to image segmentation from the results of a region growing approach using cultural algorithms," in *Proceedings of the IEEE International Conference on Evolutionary Computation*, pp. 819–824, Detroit, Mich, USA, December 1995.

[8] D. Wang, "Colony location algorithm for assignment problems," *Journal of Control Theory and Applications*, vol. 2, no. 2, pp. 111–116, 2004.

[9] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[10] B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman and Co., San Francisco, Calif, USA, 1982.

[11] H. Liu, "A fuzzy qualitative framework for connecting robot qualitative and quantitative representations," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1522–1530, 2008.

[12] R. E. Hiromoto and M. Manic, "Information-based algorithmic design of a neural network classifier," *International Scientific Journal of Computing*, vol. 5, no. 3, pp. 87–98, 2006.

[13] I. C. Trelea, "The particle swarm optimization algorithm: convergence analysis and parameter selection," *Information Processing Letters*, vol. 85, no. 6, pp. 317–325, 2003.

[14] A. Linhares, "State-space search strategies gleaned from animal behavior: a traveling salesman experiment," *Biological Cybernetics*, vol. 78, no. 3, pp. 167–173, 1998.

[15] A. Linhares, "Synthesizing a predatory search strategy for VLSI layouts," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 2, pp. 147–152, 1999.

[16] J. N. M. Smith, "The food searching behaviour of two European thrushes. II. The adaptiveness of the search patterns," *Behavior*, vol. 49, no. 1-2, pp. 1–61, 1974.

[17] K. Nakamuta, "Mechanism of the switchover from extensive to area-concentrated search behaviour of the ladybird beetle, Coccinella septempunctata bruckii," *Journal of Insect Physiology*, vol. 31, no. 11, pp. 849–856, 1985.

[18] W. J. Bell, "Searching behavior patterns in insects," *Annual Review of Entomology*, vol. 35, pp. 447–467, 1990.

[19] P. Kareiva and G. Odell, "Swarms of predators exhibit "preytaxis" if individual predators use area-restricted search," *American Naturalist*, vol. 130, no. 2, pp. 233–270, 1987.

[20] E. Curio, *The Ethology of Predation*, Springer, Berlin, Germany, 1976.

[21] C. Liu and D. Wang, "Predatory search algorithm with restriction of solution distance," *Biological Cybernetics*, vol. 92, no. 5, pp. 293–302, 2005.

[22] W. J. Zhang, *Studies on design of spatial crank-rocker mechanisms based on the pressure angle [M.S. thesis]*, Dong Hua University, Shanghai, China, 1984.

[23] J. Kennedy, "Particle swarm: social adaptation of knowledge," in *Proceedings of the IEEE International Conference on Evolutionary Computation (ICEC '97)*, pp. 303–308, Washington, DC, USA, April 1997.

[24] Y. Shi and R. Eberhart, "Modified particle swarm optimizer," in *Proceedings of the IEEE International Conference on Evolutionary Computation (ICEC '98)*, pp. 69–73, Indianapolis, Ind, USA, May 1998.

[25] J. Kennedy and R. Mendes, "Population structure and particle swarm performancein," in *Proceedings of IEEE Congress on Evolutionary Computation*, pp. 1671–1676, Honolulu, Hawaii, USA, 2002.

[26] S.-Z. Zhao and P. N. Suganthan, "Two-lbests based multiobjective particle swarm optimizer," *Engineering Optimization*, vol. 43, no. 1, pp. 1–17, 2011.

[27] C. L. Sun, J. C. Zeng, and J. S. Pan, "An improved vector particle swarm optimization for constrained optimization problems," *Information Sciences*, vol. 181, no. 6, pp. 1153–1163, 2011.

[28] R. D. Araújo, "Swarm-based translation-invariant morphological prediction method for financial time series forecasting," *Information Sciences*, vol. 180, no. 24, pp. 4784–4805, 2010.

[29] S.-K. Oh, W.-D. Kim, W. Pedrycz, and B.-J. Park, "Polynomial-based radial basis function neural networks (P-RBF NNs) realized with the aid of particle swarm optimization," *Fuzzy Sets and Systems*, vol. 163, pp. 54–77, 2011.

[30] W. Pedrycz and K. Hirota, "Fuzzy vector quantization with the particle swarm optimization: a study in fuzzy granulation-degranulation information processing," *Signal Processing*, vol. 87, no. 9, pp. 2061–2074, 2007.

[31] W. Pedrycz, B. J. Park, and N. J. Pizzi, "Identifying core sets of discriminatory features using particle swarm optimization," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4610–4616, 2009.

[32] J. W. Wang, W. H. Ip, and W. J. Zhang, "An integrated road construction and resource planning approach to the evacuation of victims from single source to multiple destinations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 277–289, 2010.

[33] J. W. Wang, H. F. Wang, W. J. Zhang, W. H. Ip, and K. Furuta, "Evacuation planning based on the contraflow technique with consideration of evacuation priorities and traffic setup time," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 480–485, 2013.

[34] J. J. Liang, A. K. Qin, P. N. Suganthan, and S. Baskar, "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 3, pp. 281–295, 2006.

[35] S. Agrawal, B. K. Panigrahi, and M. K. Tiwari, "Multiobjective particle swarm algorithm with fuzzy clustering for electrical power dispatch," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 5, pp. 529–541, 2008.

[36] Z. H. Zhan, J. Zhang, Y. Li, and H. S. H. Chung, "Adaptive particle swarm optimization," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 39, no. 6, pp. 1362–1381, 2009.

[37] M. Clerc, "Particle swarm optimization," in *International Scientific and Technical Encyclopaedia*, Wiley, Hoboken, NJ, USA, 2006.

[38] Y. Cooren, M. Clerc, and P. Siarry, "Performance evaluation of TRIBES, an adaptive particle swarm optimization algorithm," *Swarm Intelligence*, vol. 3, no. 2, pp. 149–178, 2009.

[39] J. Wang and D. Wang, "Particle swarm optimization with a leader and followers," *Progress in Natural Science*, vol. 18, no. 11, pp. 1437–1443, 2008.

[40] A. Silva, A. Neves, and E. Costa, *An Empirical Comparison of Particle Swarm and Predator Prey Optimisation*, vol. 2464 of

Lecture Notes in Computer Science, Springer, Berlin, Germany, 2002.

[41] R. Silva, A. Neves, and E. Costa, *SAPPO: A Simple, Adaptable, Predator Prey Optimiser*, vol. 2902 of *Lecture Notes in Computer Science*, Springer, Berlin, Germany, 2003.

[42] P. N. Suganthan, N. Hansen, J. J. Liang et al., "Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimizationin," in *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 1–50, 2005.

[43] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.

[44] W. Zhong, J. Liu, M. Xue, and L. Jiao, "A multiagent genetic algorithm for global numerical optimization," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 34, no. 2, pp. 1128–1141, 2004.

[45] Z. M. Bi, *Adaptive robotic manufacturing systems [Ph.D. thesis]*, Department of Mechanical Engineering, University of Saskatchewan, Saskatchewan, Canada, 2002.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 725017, 13 pages
http://dx.doi.org/10.1155/2013/725017

Research Article

Identification of Fuzzy Inference Systems by Means of a Multiobjective Opposition-Based Space Search Algorithm

Wei Huang[1,2] and Sung-Kwun Oh[3]

[1] *State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China*
[2] *School of Computer and Communication Engineering, Tianjin University of Technology, Tianjin 300384, China*
[3] *Department of Electrical Engineering, The University of Suwon, San 2-2, Wau-ri, Bongdam-eup, Gyeonggi-do, Hwaseong-si 445-743, Republic of Korea*

Correspondence should be addressed to Sung-Kwun Oh; ohsk@suwon.ac.kr

Received 2 January 2013; Revised 25 February 2013; Accepted 25 February 2013

Academic Editor: Zhuming Bi

We introduce a new category of fuzzy inference systems with the aid of a multiobjective opposition-based space search algorithm (MOSSA). The proposed MOSSA is essentially a multiobjective space search algorithm improved by using an opposition-based learning that employs a so-called opposite numbers mechanism to speed up the convergence of the optimization algorithm. In the identification of fuzzy inference system, the MOSSA is exploited to carry out the parametric identification of the fuzzy model as well as to realize its structural identification. Experimental results demonstrate the effectiveness of the proposed fuzzy models.

1. Introduction

Fuzzy modeling has been utilized in many fields for engineering, medical engineering, and even social science. Lots of diverse approaches to fuzzy modeling have been proposed in the past decades. Pioneering work by Pedrycz [1], Tong [2], Xu and Zailu [3], Sugeno and Yasukawa [4], Oh and Pedrycz [5], Chung et al. [6], and others [7] has studied different approaches to construct fuzzy models. All these methods reported are based on information granulation and optimization algorithms, and there is a lack of investigations of multiobjective identification of fuzzy models.

There have been a suite of studies focusing on multiobjective when designing fuzzy models. In the design of fuzzy models, two main and conflicting objectives are commonly considered. One is the accuracy and the other is the complexity. In the 1990s, the emphasis of modeling was positioned on the accuracy maximization. As powerful optimization tools in many science fields [8], various approaches such as Genetic Algorithms (GAs) and Particle Swarm Optimization (PSO) [9, 10] have been proposed to improve the accuracy of fuzzy models. As a result of accuracy maximization, the complexity of the model increases. Some researchers attempted to simultaneously optimize the accuracy and the complexity of the fuzzy models [11, 12]. Since it is impossible to simultaneously optimize these objectives due to the existence of the accuracy-complexity tradeoff, the accuracy maximization and complexity minimization have been often cast in the setting of multiobjective optimization. A number of evolutionary algorithms (EAs) have been developed to solve multiobjective optimization problems such as micro-GA [13] and NSGA-II [14–16]. As a result, multiobjective optimization (MOO) techniques have been applied to the design of fuzzy models exhibiting high accuracy and significant interpretability [17, 18]. Nevertheless, when dealing with the fuzzy models, the previous studies lack an optimization vehicle which considers not only the solution space being explored but also the techniques of MOO.

In our previous study [19], we propose a design of fuzzy models based on multiobjective space search algorithm (MSSA). This work provides some enhancements of the fuzzy

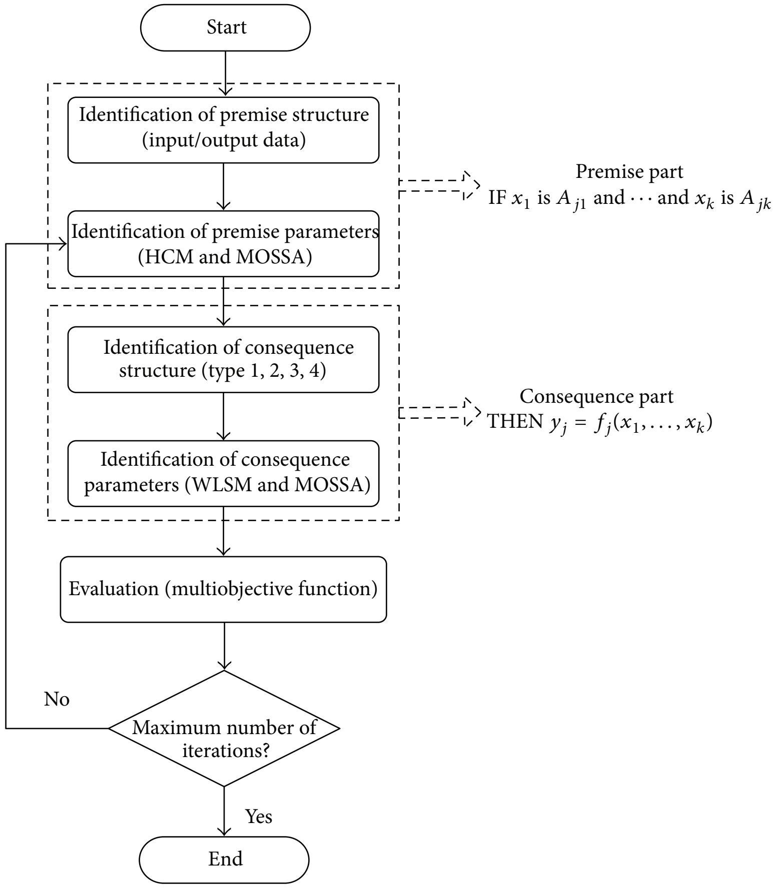


Figure 1: An overall scheme of fuzzy rule-based modeling.

modeling. Noticeably, some limitations of this fuzzy model are as follows: (1) the conventional MSSA is essentially stochastic search techniques, and such random mechanism leads to slow convergence speed; and (2) the flexibility and predictive ability are limited due to the only one single type of all polynomials of the consequence part of fuzzy rules.

In this study, we present a multiobjective opposition-based space search algorithm (MOSSA) and introduce a design of fuzzy model by means of the MOSSA and weighted least squares method (WLSM). The resulting fuzzy models address the two constrains mentioned above. First, the proposed MOSSA that is used as a vehicle to maximize the accuracy of the fuzzy inference system could come with more rapid convergence speed in comparison with the conventional MSSA. Second, instead of the ordinary least squares method (LSM), WLSM is used to estimate the coefficients of the consequent polynomials. With the use of WLSM, fuzzy model exhibits different types of polynomials, which can vary from one rule to another.

2. A Design of Fuzzy Inference Systems

Figure 1 depicts an overall scheme of fuzzy rule-based modeling. The identification procedure for fuzzy models is split into two parts, namely, premise part and consequence part of the fuzzy rules.

2.1. Identification of Premise Part. The identification completed at the premise level consists of two main steps. First, we select the input variables $x_1, x_2, \ldots, x_k$ of the rules. Second, we form fuzzy partitions (by specifying fuzzy sets of well-defined semantics e.g., Low, High, etc.) of the spaces over which these individual variables are defined.

The identification of the premise part is completed in the following manner.

Given is a set of data $\mathbf{U} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l; \mathbf{y}\}$, where $\mathbf{x}_k = [x_{1k}, \ldots, x_{mk}]^T$, $\mathbf{y} = [y_1, \ldots, y_m]^T$, where l is the number of variables and m stands for the number of data.

Step 1. Arrange a set of data $\mathbf{U}$ into data set $\mathbf{X}_k$ composed of the corresponding input and output data.

$$\mathbf{X}_k = [\mathbf{x}_k; \mathbf{y}]. \tag{1}$$

Step 2. Run the K-Means to determine the centers (prototypes) $\mathbf{v}_{kg}$ of the data set $\mathbf{X}_k$.

Step 2.1. Arrange data set $\mathbf{X}_k$ into c-clusters (in essence this is effectively the information granulation process).

Step 2.2. Calculate the centers $\mathbf{v}_{kg}$ of each cluster as follows:

$$\mathbf{v}_{kg} = \{v_{k1}, v_{k2}, \ldots, v_{kc}\}. \tag{2}$$

Step 3. Partition the corresponding input space using the prototypes of the clusters $\mathbf{v}_{kg}$. Associate each cluster with some meaning (semantics), say Small, Large, and so forth.

Step 4. Set the initial apexes of the membership functions using the prototypes $\mathbf{v}_{kg}$.

2.2. Identification of Consequence Part. The identification of the consequence part of the rules embraces two phases, namely, (1) a selection of the consequence variables of the fuzzy rules (identification of consequence structure) and (2) determination of the parameters of the consequence (identification of consequence parameters).

2.2.1. Identification of Consequence Structure. The identification of the conclusion parts of the rules deals with a selection of their structure (Types 1, 2, 3, and 4) that is followed by the determination of the respective parameters of the local functions occurring there. The consequence part of the rule that is extended form of a typical fuzzy rule in the TSK (Takagi-Sugeno-Kang) fuzzy model has the form

$$\begin{aligned} R^j &: \text{If } x_1 \text{ is } A_{1c} \text{ and } \ldots \text{ and } x_k \text{ is } A_{kc} \\ &\text{then } y_j - M_j = f_j(x_1, \ldots, x_k). \end{aligned} \tag{3}$$

Type 1 (simplified inference). Consider

$$f_j = a_{j0}. \tag{4}$$

Type 2 (linear inference). Consider

$$f_j = a_{j0} + a_{j1}(x_1 - V_{j1}) + \cdots + a_{jk}(x_k - V_{jk}). \tag{5}$$

Type 3 (quadratic inference). Consider

$$\begin{aligned} f_j = {} & a_{j0} + a_{j1}(x_1 - V_{1j}) + \cdots + a_{jk}(x_k - V_{kj}) \\ & + a_{j(k+1)}(x_1 - V_{1j})^2 + \cdots + a_{j(2k)}(x_k - V_{kj})^2 \\ & + a_{j(2k+1)}(x_1 - V_{1j})(x_2 - V_{2j}) \\ & + \cdots + a_{j((k+2)(k+1)/2)}(x_{k-1} - V_{(k-1)j})(x_k - V_{kj}). \end{aligned} \tag{6}$$

Type 4 (modified quadratic inference). Consider

$$\begin{aligned} f_j = {} & a_{j0} + a_{j1}(x_1 - V_{1j}) + \cdots + a_{jk}(x_k - V_{kj}) \\ & + a_{j(k+1)}(x_1 - V_{1j})(x_2 - V_{2j}) \\ & + \cdots + a_{j(k(k+1)/2)}(x_{k-1} - V_{(k-1)j})(x_k - V_{kj}), \end{aligned} \tag{7}$$

where R^j is the jth fuzzy rule, x_k represents the input variables, A_{kc} is a membership function of fuzzy sets, a_{jk} is a constant, V_{kj} and M_j are a center value of the input and output data, respectively, and n is the number of fuzzy rules.

The calculation of the numeric output of the model, based on the activation (matching) levels of rules there, relies on the following expression:

$$\begin{aligned} y^* &= \frac{\sum_{j=1}^{n} w_{ji} y_i}{\sum_{j=1}^{n} w_{ji}} = \frac{\sum_{j=1}^{n} w_{ji}\left(f_j(x_1, \ldots, x_k) + M_j\right)}{\sum_{j=1}^{n} w_{ji}} \\ &= \sum_{j=1}^{n} \widehat{w}_{ji}\left(f_j(x_1, \ldots, x_k) + M_j\right). \end{aligned} \tag{8}$$

Here, y^* is the inferred output value, and w_{ji} is the premise level of matching R^j (activation level).

2.2.2. Identification of Consequence Parameters. The identification of consequence parameters is completed with the aid of WLSM, which is to determine the coefficients of the model through the minimization of the objective function J_L. The main difference between the WLSM and the classical LSM is the weighting scheme, which comes as a part of the WLSM and makes its focused on the corresponding local model as follows:

$$J_L = \sum_{i=1}^{n} \sum_{k=1}^{m} w_{ik}\left(y_k - f_i(\mathbf{x}_k - \mathbf{v}_i)\right)^2. \tag{9}$$

The performance index J_L can be rearranged as

$$\begin{aligned} J_L &= \sum_{i=1}^{n} (\mathbf{Y} - \mathbf{X}_i \mathbf{a}_i)^T \mathbf{W}_i (\mathbf{Y} - \mathbf{X}_i \mathbf{a}_i) \\ &= \sum_{i=1}^{n} \left(\mathbf{W}_i^{1/2}\mathbf{Y} - \mathbf{W}_i^{1/2}\mathbf{X}_i\mathbf{a}_i\right)^T \left(\mathbf{W}_i^{1/2}\mathbf{Y} - \mathbf{W}_i^{1/2}\mathbf{X}_i\mathbf{a}_i\right), \end{aligned} \tag{10}$$

where $\mathbf{a}_i$ is the vector of coefficients of ith consequent polynomial (local model), and $\mathbf{W}_i$ is the diagonal matrix (weighting factor matrix) which involves the activation levels. $\mathbf{X}_i$ is a matrix which includes input data shifted by the locations of the information granules (more specifically, centers of clusters). In case the consequent polynomial is Type 2 (linear or a first-order polynomial), $\mathbf{X}_i$ and $\mathbf{a}_i$ are read as follows:

$$\begin{aligned} \mathbf{W}_i &= \begin{bmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{im} \end{bmatrix} \in \mathfrak{R}^{m \times m}, \\ \mathbf{X}_i &= \begin{bmatrix} 1 & (x_{11} - v_{i1}) & \cdots & (x_{l1} - v_{il}) \\ 1 & (x_{12} - v_{i1}) & \cdots & (x_{l1} - v_{il}) \\ 1 & \vdots & \ddots & \vdots \\ 1 & (x_{1m} - v_{i1}) & \cdots & (x_{lm} - v_{il}) \end{bmatrix}, \\ \mathbf{a}_i &= [a_{i0}\ a_{i1}\ \cdots\ a_{il}]. \end{aligned} \tag{11}$$

For the local learning algorithm, the objective function is defined as a linear combination of the squared error being the difference between the data and the corresponding output of each fuzzy rule, based on a weighting factor matrix. The weighting factor matrix, $\mathbf{W}_i$, captures the activation levels of input data to ith subspace. In this sense we can consider the weighting factor matrix to form a discrete version of the fuzzy linguistic representation for the corresponding subspace.

The coefficients of the consequent polynomial of the ith fuzzy rule can be determined in a usual manner, namely,

$$\mathbf{a}_i = \left(\mathbf{X}_i^T \mathbf{W}_i \mathbf{X}_i\right)^{-1} \mathbf{X}_i \mathbf{W}_i \mathbf{Y}. \tag{12}$$

Notice that the coefficients of the consequent polynomial of each fuzzy rule are computed independently using a certain subset of the training data. These computations can be implemented in parallel and in this case the overall computing load becomes unaffected by the total number of the rules.

3. Multiobjective Optimization of Fuzzy Inference Systems

Many optimization problems come with multiple objectives, which not only interact but may be in conflict. MSSA [20] is one of multiobjective optimization algorithms and it has been successfully used in the design of fuzzy models. In this study, we develop a multiobjective optimization algorithm based on an opposition-based space search algorithm as the optimization vehicle of FIS.

3.1. Multiobjective Opposition-Based Space Search Algorithm (MOSSA). We first introduce a single-objective space search algorithm (SSA), an adaptive heuristic optimization algorithm whose search method comes with the analysis of the solution space [21]. Let us recall the space search mechanism to update the current solutions. The role of space search is to generate new solutions from the old ones. The search method is based on the operator of space search, which generates two basic steps: it generates a new subspace (local area) and realizes search in this new space. The search in the new space is realized by randomly generating a new solution (individual) located in this space. Regarding the generation of the new space, we consider two cases: (a) space search based on M selected solutions (denoted here as Case I), and (b) space search based on the current best solution (Case II).

To illustrate the operator in detail, we consider the optimization problem:

$$\begin{aligned} &\min \text{(or max)} \quad f\left(x_1, x_2, \ldots, x_n\right) \\ &\text{st:} \quad x_i \in \left[l_i, u_i\right], \quad i = 1, 2, 3, \ldots, n. \end{aligned} \tag{13}$$

Here a feasible solution can be represented in the following way $(x_1, x_2, \ldots, x_n)$. Let $X^k = (x_1^k, x_2^k, \ldots, x_n^k)$ be the kth solution, $L_i = \min_{j=1}^{M} x_i^j$, $U_i = \max_{j=1}^{M} x_i^j$.

We consider two scenarios.

(a) Space search based on M selected solutions: in this case, M solutions are randomly selected from the current population. The role of this operator is to update the current solutions by new solutions approaching to the optimum. The adjacent space based on M solutions is given in the form

$$\begin{aligned} V_1 = \{&\left(x_1, x_2, \ldots, x_n\right) \mid x_i \\ &\in \left[l_i, u_i\right] \cap \left[L_i - \lambda_i \left(U_i - L_i\right),\right. \\ &\left. U_i + \lambda_i \left(U_i - L_i\right)\right], \\ &\lambda_i > 0,\ i = 1, 2, \ldots, n\}. \end{aligned} \tag{14}$$

(b) Space search based on the current best solution: in this case, the given solution is the best solution in the current population. The role of this operator is to adjust the best solution by searching the adjacent space as follows:

$$V_2 = \left\{\left(x_1, x_2, \ldots, x_i, \ldots, x_n\right) \mid x_i \in \left[F l_i, F u_i\right],\ F \in [0, 1]\right\}. \tag{15}$$

Here F is a proportion coefficient being used to adjust the size of adjacent space. In this study, F is set to 1.

To speed up the convergence of the SSA, we develop an opposition-based space search algorithm (OSSA) realized by using a mechanism of opposition-based learning (OBL) [22]. OBL has been shown to be an effective concept to enhance various optimization approaches. Let us recall the basic concept.

Opposition-Based Point [22]. Let $P = (x_1, x_2, \ldots, x_D)$ be a point in a D-dimensional space, where $x_1, x_2, \ldots, x_D \in R$ and $x_i \in [a_i, b_i]$, for all $i \in \{1, 2, \ldots, D\}$. The opposite point $\breve{P} = (\breve{x}_1, \breve{x}_2, \ldots, \breve{x}_D)$ is completely defined by its components as follows:

$$\breve{x}_i = a_i + b_i - x_i. \tag{16}$$

Opposition-Based Optimization (OBL) [22]. Let $P = (x_1, x_2, \ldots, x_D)$ be a point in a D-dimensional space (i.e., a candidate solution). Assume $f(\bullet)$ is a fitness function. According to the definition of the opposite point, we say that $\breve{P} = (\breve{x}_1, \breve{x}_2, \ldots, \breve{x}_D)$ is the opposite of site $P = (x_1, x_2, \ldots, x_D)$. Now, if $f(\breve{P}) \geq f(P)$, then the point P can be replaced with $\breve{P}$. Hence, the point and its opposite point are evaluated simultaneously in order to continue with the one of the highest fitness.

Based on this concept, we can develop the opposition-based space search operator. Assume that the current solution set (population) is P, where h is the size of solution set, P_o is the size of opposition solution set, and q is the dimension of a solution. The Pseudocode opposition-based space search can be summarized as shown in Pseudocode 1.

With understanding of the OSSA, we can develop an MOSSA. In order to improve convergence to the Pareto front as well as produce a well-distributed Pareto front, the technique of nondominated sort with the aid of the crowding distance [23] is used in the MOSSA. The details are presented in Pseudocode 2. The nondominated sort is realized with the aid of estimation of the crowding distance among solutions

Begin
  Update the dynamic interval boundaries $[a_j, b_j]$ in P.
  /∗Generate a new solution set OP by using opposition∗/
  For ($i = 1$; $i <= Po$; $i{+}{+}$)
    For ($j = 1$; $j <= q$; $j{+}{+}$)
      $OP_{i,j} = a_j + b_j - P_{i,j}$;
      IF $OP_{i,j}$ out of the box-constraint
        $x_j^{\text{new}} = \text{rand}(a_j, b_j)$;
      End IF
    End For
    Calculate the fitness value of the $OP_{i,j}$
  End For
/∗ End opposition ∗/
End

Pseudocode 1: Pseudocode of opposition-based space search.

Initialize a solution set (population);
Evaluate each solution in the solution set P;
While (not satisfy terminate condition)
  Space search (case I);
  Sort all solutions in S based on non-domination;
  Remove the worst solution from the current solution set P;
  Space search (case II);
  Sort all solutions in S based on non-domination;
  Remove the worst solution from the current solution set P;
  Complete Opposition-Based Operation, and generate solution set OP;
  Sort all solutions in $\{P, OP\}$ based on non-domination
  Select h fittest solutions from $\{P, OP\}$;
End While
Output the optimal solution.

Pseudocode 2: Pseudocode of multiobjective opposition-based space search algorithm (MOSSA).

Table 1: List of parameters of the MOSSA.

MOSSA parameters	Values in this study
Solution set (Populations)	200
Number of solutions for search space (case I)	8
Number of generations for structural optimization	$m = 150$
Number of generations of parametric optimization in each generation of structural optimization	$k = 5$
Number of generations for parametric optimization	$n = 1000$

in the current solution set. The termination condition of the MOSSA is such that all the solutions in the current population have the same fitness or terminate after a certain fixed number of generations.

3.2. Arrangement of Solutions in the MOSSA. The standard gradient-based optimization techniques might not be effective in the context of rule-based systems given their nonlinear character (in particular the form of the membership functions) and modularity of the systems. This suggests us to explore other optimization techniques. Figure 2 depicts the arrangement of solutions present in the MOSSA-based method. The first part supporting structural optimization is separated from the second part used for parametric optimization. The size of the solutions for structural optimization of the fuzzy model is determined according to the number of all input variables of the system. The size of the solutions for parametric optimization depends on structurally optimized fuzzy inference system. When running the optimization method, we use an improved tuning method.

Figure 3 illustrates the comparison of "conventional" tuning and the proposed tuning. In the "conventional" tuning, the structural and the parametric optimization are carried out sequentially. First, the structural optimization is completed with m number of generations and then we proceed with the parametric phase with n number of generations, where m and n are given numbers. The structural optimization of the fuzzy model is carried out assuming that the apexes of the membership functions are kept fixed. The fixed apexes of the membership functions are taken as the center values produced by the C-Means algorithm, while the parametric optimization is applied to the fuzzy model derived through the structural optimization. In a nutshell, from the viewpoint

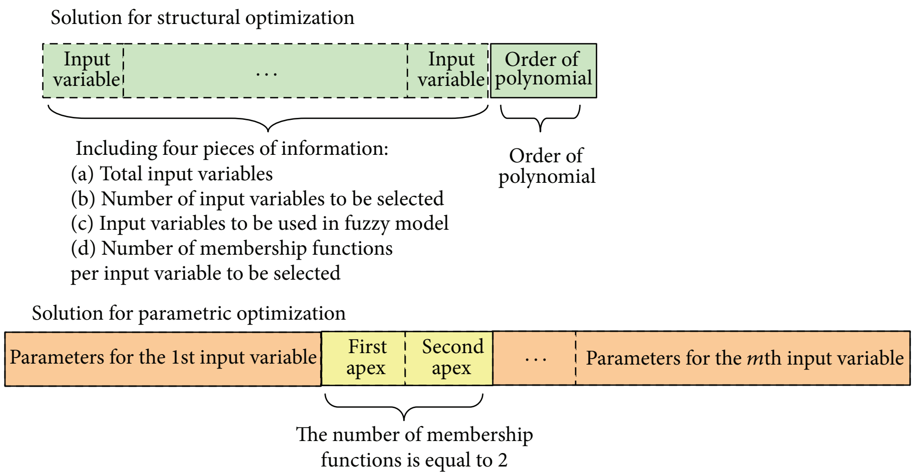


Figure 2: Arrangement of solutions using the MOSSA.

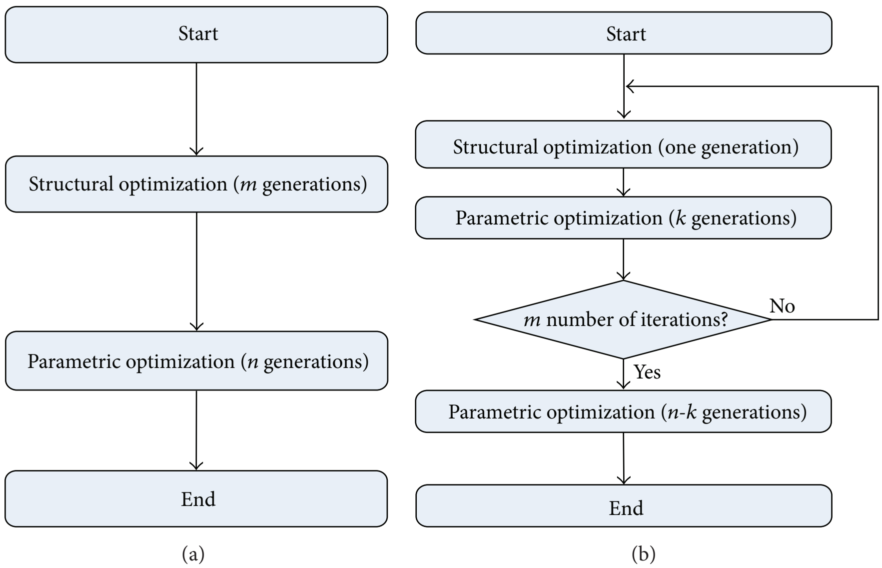


Figure 3: Comparison of "conventional" tuning and the proposed tuning: (a) overall flowchart of "conventional" tuning; and (b) overall flowchart of the proposed tuning.

of structure identification, only one fixed parameter set, which is the assigned apexes of membership functions obtained by C-Means clustering, is considered to carry out the overall structural optimization of fuzzy model. From the viewpoint of parameter identification, only one structurally optimized model that is obtained during the structure identification is considered to be involved in the overall parametric optimization. In order to construct the optimized fuzzy model, the range of search space for the structural as well as the parametric optimization is clearly restricted in the sequential tuning method. To alleviate this problem, we present a MOSSA-based improved tuning method. In this approach, we realize the structural with the aid of k number of generations of parametric optimization, where k is a given number. The several generations of parametric optimization will help to determine the optimal structure of the fuzzy model.

3.3. Objective Functions of the FIS. Three objective functions are used to evaluate the accuracy and the complexity of an FIS. Those are performance indexes, entropy of partition, and the total number of the coefficients of the polynomials to be estimated, respectively. Once the input variables of the premise part have been specified, the optimal consequence parameters that minimize the assumed performance index can be determined.

We consider two performance indexes, that is, the standard root mean squared error (RMSE) and mean squared error (MSE) as follows:

$$\text{PI (or E\_PI)} = \begin{cases} \sqrt{\dfrac{1}{m}\displaystyle\sum_{i=1}^{m}\left(y_i - y_i^*\right)^2}, & \text{(RMSE)}, \\ \dfrac{1}{m}\displaystyle\sum_{i=1}^{m}\left(y_i - y_i^*\right)^2, & \text{(MSE)}, \end{cases} \tag{17}$$

TABLE 2: Optimal solutions based on training data and testing data (NO_x).

(a) Based on the LSM

Number of solutions	Selected input variables	Number of MFs	Type	PI	E_PI	Objective values		
						E	H	N
1	**1,4, 5**	**2, 2, 2**	**3**	**0.042**	**0.104**	**0.073**	**8**	**80**
2	1, 4, 5	2, 2, 2	4	0.087	0.197	0.142	8	56
3	1, 4, 5	2, 2, 2	2	0.371	0.412	0.392	8	32
4	1, 4	2, 2	3	1.016	0.810	0.913	4	24
5	1, 2, 4, 5	2, 2, 2, 2	1	2.823	6.997	4.910	16	16
6	4, 5	2, 2	4	5.525	8.527	7.026	4	16
7	4, 5	2, 2	2	6.029	10.41	8.217	4	12
8	2, 4, 5	2, 2, 2	1	6.290	13.74	10.02	8	8
9	4, 5	2, 2	1	13.60	20.23	16.92	4	4

(b) Based on the WLSM

Number of solutions	Selected input variables	Number of MFs	Type	PI	E_PI	Objective values		
						E	H	N
1	**1, 2, 3, 4, 5**	**2, 2, 2, 2, 2**	**4**	**0.00085**	**0.00042**	**0.00064**	**32**	**512**
2	1, 2, 4, 5	2, 2, 2, 2	4	0.022	0.036	0.029	16	176
3	1, 4, 5	2, 2, 2	4	0.096	0.14	0.12	8	56
4	1, 3, 5	2, 2, 2	2	3.11	2.07	2.59	8	32
5	4, 5	2, 2	4	5.69	6.80	6.24	4	16
6	4, 5	2, 2	2	7.55	9.20	8.38	4	12
7	2, 4, 5	2, 2, 2	1	39.45	32.55	36.01	8	8
8	2, 5	2, 2	1	57.86	58.86	58.36	4	4

where y^* is the output of the fuzzy model, m is the total number of data, and i is the data index.

The accuracy criterion E includes both the training data and testing data and comes as a convex combination of the two components as follows:

$$E = \theta \times \mathrm{PI} + (1 - \theta) \times \mathrm{E\_PI}. \tag{18}$$

Here, PI and E_PI (V_PI) denote the performance index for the training data and testing data (Validation data), respectively. θ is a weighting factor that allows us to strike a sound balance between the performance of the model for the training and testing data. Depending upon the values of the weighting factor, several specific cases of the objective function are worth distinguishing.

(i) If $\theta = 1$, then the model is optimized based on the training data. No testing data is taken into consideration.

(ii) If $\theta = 0.5$, then both the training and testing data are taken into account. Moreover it is assumed that they exhibit the same impact on the performance of the model.

(iii) The case $\theta = \alpha$ where $\alpha \in [0, 1]$ embraces both the cases stated above. The choice of α establishes a certain tradeoff between the approximation and generalization aspects of the fuzzy model.

As a measure for evaluating the structure complexity of a model we consider the following partition criterion:

$$H = \prod_{i=1}^{n} F_i, \tag{19}$$

where n is the total number of selected input variables, and F_i is the number of membership functions for the ith corresponding input variable.

As a simplicity criterion we consider the consequence part of the local models, which is computed as

$$N = \sum_{j=1}^{n} C_i,$$

$$C_i = \begin{cases} 1 & \text{if type of local model is constant} \\ 1 + l & \text{if type of local model is linear form} \\ 1 + l + \dfrac{(l^2 - l)}{2} + l & \text{if type of local model is quadratic form} \\ 1 + l + \dfrac{(l^2 - l)}{2} & \text{if type of local model is modified quadratic form,} \end{cases} \tag{20}$$

TABLE 3: Optimal solutions based on training data, validation data, and testing data (NO_x).

(a) Based on LSM

Number of solutions	Selected input variables	Number of MFs	Type	PI	V_PI	E_PI	Objective values E	H	N
1	**1, 4, 5**	**2, 2, 2**	**3**	**0.014**	**0.044**	**7.046**	**0.029**	**8**	**80**
2	1, 4, 5	2, 2, 2	4	0.053	0.062	146.31	0.057	8	56
3	1, 4, 5	2, 2, 2	2	0.091	0.118	117.79	0.104	8	32
4	1, 4	2, 2	3	1.149	0.447	3.503	0.798	4	24
5	4, 5	2, 2	4	5.269	5.409	16.40	5.339	4	16
6	4, 5	2, 2	2	5.985	5.789	21.94	5.887	4	12
7	1, 4, 5	2, 2, 2	1	9.097	8.655	15.80	8.876	8	8
8	4, 5	2, 2	1	14.10	18.95	20.64	16.52	4	4

(b) Based on WLSM

Number of solutions	Selected input variables	Number of MFs	Type	PI	V_PI	E_PI	Objective values E	H	N
1	1, 2, 3, 4, 5	2, 2, 2, 2, 2	4	0.00068	0.00062	0.0000074	0.00065	32	512
2	**1, 3, 4, 5**	**2, 2, 2, 2**	**4**	**0.0086**	**0.0068**	**0.0804**	**0.00765**	**16**	**176**
3	1, 4, 5	2, 2, 2, 2, 2	4	0.097	0.122	0.135	0.109	8	56
4	1, 3, 5	2, 2, 2, 2	2	2.55	0.596	2.949	1.572	8	32
5	1, 4	2, 2	4	5.88	1.456	3.419	3.667	4	16
6	4, 5	2, 2	2	7.51	5.255	12.39	6.383	4	12
7	2, 4, 5	2, 2, 2	1	39.53	38.74	19.03	3.914	8	8
8	4, 5	2, 2	1	52.44	40.92	39.90	4.668	4	4

TABLE 4: Comparative analysis of selected models (NO_x).

Model			PI (MSE)	V_PI (MSE)	E_PI (MSE)	Number of rules
Regression model			17.68		19.23	
Hybrid FS-FNNs [20]			2.806		5.164	30
Hybrid FR-FNNs [24]			0.080		0.190	32
Multi-FNN [25]			0.720		2.205	30
Hybrid rule-based FNNs [26]			3.725		5.291	30
Choi's model [27]			0.012		0.067	18
Our model	LSM	Two split dataset	0.042		0.104	8
		Three split dataset	0.014	0.044	7.046	8
	WLSM	Two split dataset	0.022		0.036	16
		Three split dataset	0.009	0.007	0.080	16

where C_i is the number of coefficients of the ith polynomial and l stands for the number of input variables.

In a nutshell, we find the Pareto optimal sets and Pareto front by minimizing $\{E, H, N\}$ by means of the MOSSA. This leads to easily interpretable, simple, and accurate fuzzy models.

4. Experimental Studies

This section reports on comprehensive numeric studies illustrating the design of the fuzzy model and quantifying its performance. We use three well-known data sets. Each data set is divided into two parts of the same size. PI denotes the performance index for the training data, V_PI represents the validation data, and E_PI stands for the testing data. In all considerations, the weighting factor θ was set to 0.5.

The parameters of the MOSSA are summarized in Table 1 (the choice of these specific values of the parameters is a result of intensive experimentations; as a matter of fact, those values are in line with those reported in the literature).

4.1. NO_x Emission Process Data of Gas Turbine Power Plant. NO_x emission process is modeled using the data of gas turbine power plants. A NO_x emission process of a GE gas turbine power plant located in Virginia, USA, is chosen in this experiment. The input variables include AT (ambient temperature a site), CS (compressor speed), LPTS (low pressure turbine speed), CDP (compressor discharge pressure), and

Table 5: Optimal solutions based on training data and testing data (MPG).

(a) Based on the LSM

Number of solutions	Selected input variables	Number of MFs	Type	PI	E_PI	Objective values		
						E	H	N
1	**4, 5, 6**	**2, 2, 2**	**2**	**2.535**	**2.372**	**2.454**	**8**	**32**
2	4, 6	2, 2	2	2.792	2.718	2.755	4	12
3	4, 5, 6	2, 2, 2	1	2.893	2.585	2.739	8	8
4	4, 6	2, 2	1	2.982	2.728	2.855	4	4

(b) Based on the WLSM

Number of solutions	Selected input variables	Number of MFs	Type	PI	E_PI	Objective values		
						E	H	N
1	1, 2, 3, 5, 6, 7	2, 2, 2, 2, 2, 2	4	0.646	0.245	0.445	64	1408
2	**3, 4, 5, 6, 7**	**2, 2, 2, 2, 2**	**4**	**1.351**	**0.683**	**1.017**	**32**	**512**
3	2, 5, 6, 7	2, 2, 2, 2	4	1.551	0.940	1.246	16	176
4	1, 4, 5, 6	2, 2, 2, 2	2	2.266	1.689	1.977	16	80
5	2, 5, 6	2, 2, 2	4	2.396	1.658	2.027	8	56
6	**4, 5, 6**	**2, 2, 2**	**2**	**2.636**	**2.132**	**2.384**	**8**	**32**
7	4, 6	2, 2	4	2.849	2.399	2.624	4	16
8	4, 6	2, 2	2	2.889	2.457	2.673	4	12
9	2, 3, 6	2, 2, 2	1	3.228	2.668	2.948	8	8
10	4, 6	2, 2	1	3.355	2.982	3.168	4	4

Table 6: Optimal solutions based on training data, validation data, and testing data (MPG).

(a) Based on the LSM

Number of solutions	Selected input variables	Number of MFs	Type	PI	V_PI	E_PI	Objective values		
							E	H	N
1	2, 3, 6	2, 2, 2	4	2.284	2.632	3.174	2.458	8	56
2	**3, 4, 6**	**2, 2, 2**	**2**	**2.450**	**2.757**	**2.536**	**2.603**	**8**	**32**
3	3, 4, 5, 6	2, 2, 2, 2	1	2.679	2.704	2.292	2.692	16	16
4	4, 6	2, 2	2	2.781	2.819	2.592	2.800	4	12
5	3, 4, 6	2, 2, 2	1	2.772	2.895	2.290	2.834	8	8
6	4, 6	2, 2	1	3.037	2.792	2.773	2.914	4	4

(b) Based on the WLSM

Number of solutions	Selected input variables	Number of MFs	Type	PI	V_PI	E_PI	Objective values		
							E	H	N
1	2, 3, 4, 5, 6, 7	2, 2, 2, 2, 2, 2	4	0.451	0.083	0.096	0.267	64	1408
2	**2, 3, 5, 6, 7**	**2, 2, 2, 2, 2**	**4**	**0.637**	**0.088**	**0.199**	**0.362**	**32**	**512**
3	1, 2, 4, 5, 6, 7	2, 2, 2, 2, 2, 2	2	1.589	1.044	0.897	1.316	64	448
4	3, 5, 6, 7	2, 2, 2, 2	4	1.802	0.842	0.978	1.322	16	176
5	1, 4, 5, 6	2, 2, 2, 2	2	2.206	1.658	1.719	1.932	16	80
6	**2, 5, 6**	**2, 2, 2**	**4**	**2.368**	**1.618**	**1.282**	**1.993**	**8**	**56**
7	3, 4, 6	2, 2, 2	2	2.548	2.284	1.708	2.416	8	32
8	4, 6	2, 2, 2	4	2.876	2.532	2.026	2.704	4	16
9	4, 6	2, 2	2	2.913	2.634	2.350	2.773	4	12
10	3, 4, 6	2, 2, 2	1	2.974	2.887	2.763	2.930	8	8
11	4, 6	2, 2	1	3.423	3.036	3.174	3.230	4	4

TET (turbine exhaust temperature). The output variable is NO_x. The performance index is MSE defined by (17).

First, the NO_x emission process is split into two separate data sets. The first 50% of data set (consisting of 130 pairs) is used in the design of the fuzzy model. The remaining 50% data set (consisting of 130 pairs) helps quantify the predictive quality of the model. Table 2 summarizes the performance values of optimal solutions (individuals) objective functions

Table 7: Comparative analysis of the selected models (MPG).

Model			PI (RMSE)	V_PI (RMSE)	E_PI (RMSE)	Number of rules
RBFNN [28]			3.24		3.62	36
Linguistic model [10]			2.86		3.24	36
Functional RBFNN [28]			2.41		2.82	33
Our model	LSM	Two split dataset	2.54		2.37	8
		Three split dataset	2.45	2.76	2.54	8
	WLSM	Two split dataset	2.40		1.66	8
			1.36		0.68	32
		Three split dataset	2.37	1.62	1.28	8
			0.64	0.09	0.20	32

Table 8: Optimal solutions based on training data and testing data (Housing).

(a) Based on the LSM

Number of solutions	Selected input variables	Number of MFs	Type	PI	E_PI	Objective values		
						E	H	N
1	6, 9, 10, 11, 13	2, 2, 2, 2, 2	2	3.148	3.306	3.227	32	192
2	**6, 7, 10, 11, 13**	**2, 2, 2, 2, 2**	**1**	**3.459**	**3.637**	**3.548**	**32**	**32**
3	1, 5, 6	2, 2, 2	4	4.275	4.171	4.223	8	56
4	3, 6, 13	2, 2, 2	1	4.114	3.907	4.008	8	8
5	2, 3	2, 2	3	7.087	7.593	7.340	4	24
6	3, 11	2, 2	1	7.300	7.496	7.398	4	4

(b) Based on the WLSM

Number of solutions	Selected input variables	Number of MFs	Type	PI	E_PI	Objective values		
						E	H	N
1	3, 6, 7, 8, 11, 13	2, 2, 2, 2, 2, 2	4	1.208	0.077	0.643	64	1408
2	1, 5, 6, 8, 13	2, 2, 2, 2, 2	4	1.319	0.726	1.022	32	512
3	6, 7, 10, 11, 12, 13	2, 2, 2, 2, 2, 2	2	2.134	1.410	1.772	64	448
4	1, 6, 9, 11, 13	2, 2, 2, 2, 2	2	2.776	2.105	2.440	32	192
5	**1, 6, 13**	**2, 2, 2**	**4**	**2.872**	**2.272**	**2.572**	**8**	**56**
6	6, 7, 10, 12, 13	2, 2, 2, 2, 2	1	3.939	3.687	3.813	32	32
7	2, 6, 10	2, 2, 2	2	5.533	3.893	4.713	8	32
8	6, 11, 13	2, 2, 2	1	4.776	3.970	4.373	8	8
9	3, 13	2, 2	1	5.824	5.961	5.892	4	4

(E, H, N) of the FIS. Generally, as could have been expected, by increasing the number of coefficients or rules, the accuracy of FIS becomes better.

Second, the NO_x emission process data is divided into three parts. 130 pairs out of 260 pairs of input-output data are used for training; 78 pairs out of 260 pairs of input-output data are utilized for validation; the remaining part (consisting of 52 pairs) serves as a testing set. The results of the proposed model are shown in Table 3.

Table 4 illustrates the results of comparative analysis of the proposed model when being contrasted with other models. The selected values of the performance indexes of the FIS are included in Tables 2 and 3, respectively. We note that the proposed model outperforms several previous fuzzy models known in the literature.

4.2. Automobile Miles Per Gallon (MPG) Data. We consider automobile MPG data (http://archive.ics.uci.edu/ml/datasets/Auto+MPG) with the output being the automobile's fuel consumption expressed in miles per gallon. The data set includes 392 input-output pairs (after removing incomplete instances) where the input space involves 8 input variables. To come up with a quantitative evaluation of the fuzzy model, we use the standard RMSE performance index as the one described by (17).

The automobile MPG data is partitioned into two separate parts. The first 235 data pairs are used as the training data set for FIS while the remaining 157 pairs are the testing data set for assessing the predictive performance. Table 5 summarizes the performance values of solution (individual) objective functions (E, H, N) of the FIS.

Table 9: Optimal solutions based on training data, validation data, and testing data (Housing).

(a) Based on the LSM

Number of solutions	Selected input variables	Number of MFs	Type	PI	V_PI	E_PI	Objective values		
							E	H	N
1	**1, 6, 11, 13**	**2, 2, 2, 2**	**1**	**3.529**	**2.888**	**5.093**	**3.209**	**16**	**16**
2	1, 6, 13	2, 2, 2	1	3.680	3.312	5.713	3.496	8	8
3	6, 13	2, 2	1	4.662	3.299	4.859	3.980	4	4

(b) Based on the WLSM

Number of solutions	Selected input variables	Number of MFs	Type	PI	V_PI	E_PI	Objective values		
							E	H	N
1	3, 6, 8, 11, 12, 13	2, 2, 2, 2, 2, 2	4	0.850	0.031	0.163	0.441	64	1408
2	5, 6, 8, 11, 13	2, 2, 2, 2, 2	4	1.087	0.448	0.510	0.768	32	512
3	**6, 8, 11, 13**	**2, 2, 2, 2**	**4**	**1.652**	**1.210**	**1.738**	**1.431**	**16**	**176**
4	1, 6, 13	2, 2, 2	4	2.841	2.321	2.246	2.581	8	56
5	3, 6, 7, 9, 13	2, 2, 2, 2, 2	1	3.965	3.300	4.678	3.633	32	32
6	4, 6, 13	2, 2, 2	1	5.042	3.487	4.623	4.264	8	8
7	10, 13	2, 2	1	5.720	5.347	7.020	5.533	4	4

Table 10: Comparative analysis of selected models (Boston Housing).

Model			PI (RMSE)	V_PI (RMSE)	E_PI (RMSE)	Number of rules
FIS [27]			3.13		3.17	12
RBFNN [28]			6.63		7.14	20
SVR [29]			1.17		5.84	
MARS [30]			0.97		5.96	
NN			4.19		4.26	12
			3.27		5.14	24
FNN [31]			3.76		4.08	21
FPNN			3.51		16.93	16
Our model	LSM	Two split dataset	3.46		3.64	32
		Three split dataset	3.53	2.89	5.09	16
	WLSM	Two split dataset	2.87		2.27	8
		Three split dataset	1.65	1.21	1.73	16

Next, we divide the automobile MPG data into three separate parts. The first one (consisting of 196 pairs) is used for training. The second (consisting of 118 pairs) part of the series is utilized for validation. The remaining part (consisting of 78 pairs) serves as a testing set. The values of the performance index are presented in Table 6.

The identification error of the proposed model is compared with the performance of some other model; refer to Table 7. The selected values of the performance indexes of the FIS are marked in Tables 5 and 6, respectively. The performance of the proposed model is better in the sense of its approximation and prediction abilities.

4.3. Boston Housing Data. Here we experiment with the Boston housing data set [29]. This data set concerns a description of real estate in the Boston area where houses are characterized by features such as crime rate, size of lots, number of rooms, age of houses, and their median price. The dataset consists of 506 14-dimensional data. The performance index is defined as the RMSE as given by (17).

We consider the Boston Housing data set, which is split into two separate parts. The construction of the fuzzy model is completed for 253 data points being regarded as a training set. The rest of the data set (i.e., 253 data points) is retained for testing purposes. The values of the performance index are summarized in Table 8.

Next, we move on to the Boston Housing data set, which is partitioned into three separate parts. 253 pairs out of 506 pairs of input-output data are used for training; 152 pairs out of 506 pairs of input-output data are utilized for validation; the remaining part (consisting of 101 pairs) serves as a testing set. Table 9 summarizes the results obtained for the optimized structure and performance index for optimized parameters by the MOSSA.

Table 10 supports a comparative analysis considering some existing models. The selected values of the performance indexes of the FIS are marked in Tables 8 and 9, respectively. It is evident that the proposed model compares favorably both in terms of accuracy and prediction capabilities.

5. Concluding Remarks

This paper contributes to the research area of the hybrid optimization of fuzzy inference systems in the following two important aspects: (1) we proposed a multiobjective oppisiton-based space search algorithm; (2) we introduced the identification of fuzzy inference systems based on the MOSSA and WLSM. Numerical experiments using three well-known dataset show that the model constructed with the aid of the MOSSA exhibits better performance in comparison with the fuzzy model reported in the literature.

Acknowledgments

This work was supported by the Open Foundation of State Key Laboratory of Virtual Reality and Technology and Systems, Beihang University, China (Grant no. BUAA-VR-12KF-13), by Educational Commission of Tianjin City, China (Grant no. 20120703), and also supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2012-003568). This work was partly supported by the GRRC program of Gyeonggi province (GRRC Suwon 2012-B2, Center for U-city Security & Surveillance Technology).

References

[1] W. Pedrycz, “An identification algorithm in fuzzy relational systems,” *Fuzzy Sets and Systems*, vol. 13, no. 2, pp. 153–167, 1984.

[2] R. M. Tong, “The evaluation of fuzzy models derived from experimental data,” *Fuzzy Sets and Systems*, vol. 4, no. 1, pp. 1–12, 1980.

[3] C. W. Xu and Y. Zailu, “Fuzzy model identification self-learning for dynamic system,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 17, no. 4, pp. 683–689, 1987.

[4] M. Sugeno and T. Yasukawa, “Linguistic modeling based on numerical data,” in *Proceedings of the International Fuzzy Systems Association Congress (IFSA ’91)*, Computer, Management & System Science, pp. 264–267, Brussels, Belgium, 1991.

[5] S. Oh and W. Pedrycz, “Identification of fuzzy systems by means of an auto-tuning algorithm and its application to nonlinear systems,” *Fuzzy Sets and Systems*, vol. 115, no. 2, pp. 205–230, 2000.

[6] Y. Chung, W. Pedrycz, and E. T. Kim, “A new two-phase approach to fuzzy modeling for nonlinear function approximation,” *IEICE Transactions on Information and Systems*, vol. E89-D, no. 9, pp. 2473–2483, 2006.

[7] R. Alcalá, J. Alcalá-Fdez, F. Herrera, and J. Otero, “Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation,” *International Journal of Approximate Reasoning*, vol. 44, no. 1, pp. 45–64, 2007.

[8] W. Huang and L. Ding, “Project-scheduling problem with random time-dependent activity duration times,” *IEEE Transactions on Engineering Management*, vol. 58, no. 2, pp. 377–387, 2011.

[9] A. Fernández, M. J. del Jesus, and F. Herrera, “Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets,” *International Journal of Approximate Reasoning*, vol. 50, no. 3, pp. 561–577, 2009.

[10] M. P. Ana, S. Luciano, and C. Ines, “Diagnosis of dyslexia with low quality data with genetic fuzzy systems,” *International Journal of Approximate Reasoning*, vol. 51, no. 8, pp. 993–1009, 2010.

[11] Y. Jin, “Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement,” *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 2, pp. 212–221, 2000.

[12] M. Setnes and H. Roubos, “GA-based modeling and classification: complexity and performance,” *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 5, pp. 509–522, 2000.

[13] C. Coello and G. Pulido, “Multiobjective optimization using a micro-genetic algorithm,” in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO ’01)*, pp. 274–282, 2001.

[14] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[15] G. Avigad and A. Moshaiov, “Interactive evolutionary multiobjective search and optimization of set-based concepts,” *IEEE Transactions on Systems, Man and Cybernetics B*, vol. 38, no. 2, pp. 381–403, 2008.

[16] G. G. Yen and W. F. Leong, “Dynamic multiple swarms in multiobjective particle swarm optimization,” *IEEE Transactions on Systems, Man, and Cybernetics A*, vol. 39, no. 4, pp. 890–911, 2009.

[17] L. J. Herrera, H. Pomares, I. Rojas, O. Valenzuela, and A. Prieto, “TaSe, a Taylor series-based fuzzy system model that combines interpretability and accuracy,” *Fuzzy Sets and Systems*, vol. 153, no. 3, pp. 403–427, 2005.

[18] M. Delgado, M. P. Ceullar, and M. C. Pegalajar, “Multiobjective hybrid optimization and training of recurrent neural networks,” *IEEE Transactions on Systems, Man and Cybernetics B*, vol. 38, no. 2, pp. 381–403, 2008.

[19] W. Huang, L. Ding, S. K. Oh, L. Ding, H. K. Kim, and S. C. Joo, “Identification of fuzzy inference systems using a multiobjective space search algorithm and information granulation,” *Journal of Electrical Engineering & Technology*, vol. 6, no. 6, pp. 853–866, 2011.

[20] S.-K. Oh, W. Pedrycz, and H.-S. Park, “Hybrid identification in fuzzy-neural networks,” *Fuzzy Sets and Systems*, vol. 138, no. 2, pp. 399–426, 2003.

[21] W. Huang, L. Ding, S. K. Oh, C. W. Jeong, and S. C. Joo, “Identification of fuzzy inference system based on information granulation,” *KSII Transactions on Internet and Information Systems*, vol. 4, no. 4, pp. 575–594, 2010.

[22] R. S. Rahnamayan, H. R. Tizhoosh, and M. M. A. Salama, “Opposition-based differential evolution,” *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 1, pp. 64–79, 2008.

[23] I. Karahan and M. Koksalan, “A territory defining multiobjective evolutionary algorithms and preference incorporation,” *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 4, pp. 636–664, 2012.

[24] H. S. Park and S. K. Oh, “Fuzzy relation-based fuzzy neural-networks using a hybrid identification algorithm,” *International Journal of Control, Automation and Systems*, vol. 1, no. 2, pp. 289–300, 2003.

[25] H. S. Park and S. K. Oh, "Multi-FNN identification based on HCM clustering and evolutionary fuzzy granulation," *International Journal of Control, Automation and Systems*, vol. 1, no. 2, pp. 194–202, 2003.

[26] S. K. Oh, W. Pedrycz, and H. S. Prak, "Implicit rule-based fuzzy-neural networks using the identification algorithm of hybrid scheme based on information granulation," *Advanced Engineering Informatics*, vol. 16, no. 4, pp. 247–263, 2002.

[27] J. N. Choi, S. K. Oh, and W. Pedrycz, "Identification of fuzzy relation models using hierarchical fair competition-based parallel genetic algorithms and information granulation," *Applied Mathematical Modelling*, vol. 33, no. 6, pp. 2791–2807, 2009.

[28] W. Pedrycz, H. S. Park, and S. K. Oh, "A granular-oriented development of functional radial basis function neural networks," *Neurocomputing*, vol. 72, no. 1–3, pp. 420–435, 2008.

[29] H. D. Chris, J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," *Advances in Neural Information Processing Systems*, vol. 9, pp. 155–161, 1997.

[30] J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991.

[31] A. F. Gobi and W. Pedrycz, "Fuzzy modelling through logic optimization," *International Journal of Approximate Reasoning*, vol. 45, no. 3, pp. 488–510, 2007.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 607829, 7 pages
http://dx.doi.org/10.1155/2013/607829

Research Article

A New Representation of Efficient Point Sets and Its Applications in DEA

Maoqin Li

Department of Mathematics, Taizhou University, Linhai, Zhejiang 317000, China

Correspondence should be addressed to Maoqin Li; limaoqin@hotmail.com

Received 18 January 2013; Accepted 25 February 2013

Academic Editor: Zhuming Bi

$E(M, K)$ ($E_w(M, K)$), the set of Pareto efficient (weak efficient) points of a set M with respect to a cone K in R^n, is expressed as a differencebetween two sets M and $M + K\backslash\{0\}$ (M and $M + \text{int}K$). Using the new representation, the properties of $E(M, K)$ are proved more easily than before. When M or K is in the form of union, intersection, sum, or difference of two sets or two cones, respectively, the properties of $E(M, K)$ are considered. Most of the properties are proved by the binary operations of sets, which is a new method in the multiobjective optimization. Then these properties are used to solve some types of multiobjective linear programming problems corresponding to Data Envelopment Analysis (DEA) models. The structures of the DEA efficient solution sets of four most representative DEA models are developed. Further more, the relationships between efficiencies of the four DEA models are deduced.

1. Introduction and Preliminaries

Multiobjective programming is the process of simultaneously optimizing two or more conflicting objectives subject to certain constraints (see [1–3]). Data Envelopment Analysis (DEA) is a nonparametric method in operations research and economics for the estimation of production frontiers. It is used to empirically measure productive efficiency of decision maker units $\text{DMU}_1, \ldots, \text{DMU}_n$ by solving the linear programming [4–6]. Charnes et al. and Wei et al. establish the equivalence of (weak) DEA efficient solutions in DEA model and (weak) Pareto solutions of multiobjective linear programming [7–9]. There is a multiobjective linear programming, corresponding to a DEA model, such that a DMU_{j_0} ($1 \leq j_0 \leq n$) is (weak) DEA efficient if and only if (x_{j_0}, y_{j_0}) (associating with the DMU_{j_0}) is a (weak) Pareto efficient solution of the multiobjective linear programming whose feasible region is the production possibility set (see [10]).

In this paper, we propose a new representation for the set of Pareto efficient (weak efficient) points. With the help of the new representation, not only the properties of the set of Pareto efficient (weak efficient) points which are given in [1, 3, 11] can be proved more simply, but also more new properties can be obtained. It is these new properties that reveal the relationships between the set of solutions and different multiobjective linear programmings which correspond to different DEA models. Further, the relationships between efficiencies of DMUs in different DEA models are obtained by a new way. Wei et al. [10] develop a famous method to translate production possibility sets in the intersection form and in the sum form and find all DEA efficient DMUs. For each of the four most representative DEA models, we offers a simple way to get all DEA efficient DMUs by the binary operations of sets.

Now let us recall the definition of efficiency and the representation of the set of efficient points deduced by the definition in vector optimization.

Definition 1 ([1], efficiency, weak efficiency). Given a nonempty set M and a cone K with $\text{int}\,K \neq \emptyset$ in R^n, $x_0 \in M$ is called a Pareto efficient (weak efficient) point of M, if there is no $y \in M$ with $y \neq x_0$ such that $x_0 \in y + K$ ($x_0 \in y + \text{int}\,K$). The set of all Pareto efficient (weak efficient) points of M is denoted by $E(M, K)$ ($E_w(M, K)$).

$E(M,K)$ ($E_w(M,K)$) is called the efficient (weak efficient) point set of M. By Definition 1, we have that

$$E(M,K) = \{x \in M \mid \text{there is no } y \in M, \; y \neq x \text{ such that } x \in y + K\}, \tag{1}$$

$$(E_w(M,K) = \{x \in M \mid \text{there is no } y \in M \text{ such that } x \in y + \text{int } K\}). \tag{2}$$

In [2] when $K = R^n_+$, the set is also described as follows:

$$E(M,K) = \{x \in M \mid M \cap (x - K) = \{x\}\}, \tag{3}$$

$$(E_w(M,K) = \{x \in M \mid M \cap (x - \text{int } K) = \emptyset\}). \tag{4}$$

Since $K = R^n_+$ is a convex pointed cone, (1) and (3) are equivalent. So are (2) and (4). In Section 2, we give a new representation of the efficient (weak efficient) point set, which is expressed as the difference of two sets. The idea of the new representation is motivated by the following facts in the area of DEA:

(i) the structures of the production possibility sets and the relationships between these sets (5);

(ii) the equivalence of Pareto efficiency in multiobjective linear programming and DEA efficiency in DEA model (Theorem 18);

(iii) the particularity of structures of the set of solutions to the multiobjective linear programmings corresponding to DEA models (detailed in Section 3).

For the four most representative DEA models C^2R, BC^2, FG, and ST (for the details about the models, see [4, 5]), each of the DEA models associates with a production possibility set which is also the feasible set of the multiobjective linear programming corresponding to this DEA model. The production possibility sets are denoted by T_{C^2R}, T_{BC^2}, T_{FG}, and, T_{ST}, respectively. The following relations hold (the structures of these production possibility sets are presented in Section 3):

$$\begin{aligned} &\text{(a) } T_{C^2R} = T_{FG} \cup T_{ST}, \qquad &&\text{(b) } T_{BC^2} = T_{FG} \cap T_{ST}, \\ &\text{(c) } T_{C^2R} \supset T_{FG} \supset T_{BC^2}, \qquad &&\text{(d) } T_{C^2R} \supset T_{ST} \supset T_{BC^2}. \end{aligned} \tag{5}$$

It is the specialty of the relations of the production possibility sets and the equivalence of DEA efficiency and Pareto efficiency that motivate us to propose a new representation of (weak) efficient point set. Using the new representation,we obtain some new properties of efficient point set $E(M,K)$, when the set M or K is in form at union, intersection, sum, or difference of two sets. By these properties, it is easier to get the relationship between the DEA efficiency of the four DEA models than before.

This paper is organized as follows. Section 2 introduces the new representation of the efficient (weak efficient) point set of a set, discusses some new properties of efficient point set. Using the new expression of the set $E(M,K)$ ($E_w(M,K)$), most of these properties are proved by the binary operations of sets. The multiobjective linear programming problems corresponding to the four DEA models are studied in Section 3. The structures of the efficient point sets and the efficient solution sets of the multiobjective linear programming problems are developed, and then the relationships between DEA efficiencies of DMUs in four DEA models are revealed. Section 4 is devoted to the conclusion.

The following notations are used in the paper.

Let M, M_1, M_2, and K be sets in R^n:

$$\begin{aligned} M_1 + M_2 &= \{x_1 + x_2 \mid x_1 \in M_1, x_2 \in M_2\}, \\ M_1 \setminus M_2 &= \{x \mid x \in M_1, x \notin M_2\}, \\ \overline{M} &= \{x \in R^n \mid x \notin M\}, \\ K_0 &= K \setminus \{0\}. \end{aligned} \tag{6}$$

2. Some Properties of the Efficient Point Set

In this section, a new representation of $E(M,K)$ ($E_w(M,K)$) is presented. Then we prove that it is equivalent to the original ones when K is a cone. Lastly we focus on the properties of $E(M,K)$, when M or K is in the form of the union, intersection, sum, or difference of two sets. Most of the proofs are completed by the binary operations of sets, which is a new method in multiobjective optimization.

Definition 2. Given a nonempty set M and a cone K with int $K \neq \emptyset$ in R^n, the efficient (weak efficient) point set of M with respect to K is defined by

$$E(M,K) = M \setminus (M + K_0), \tag{7}$$

$$(E_w(M,K) = M \setminus (M + \text{int } K)). \tag{8}$$

Clearly, the following result holds.

Theorem 3. *Equations* (1) *and* (7) *((*2*) and* (8)*) are equivalent, that is,*

$$\begin{aligned} E(M,K) &= \{x \in M \textit{ there is no } y \in M, y \neq x \textit{ such that } x \in y + K\} = M \setminus (M + K_0), \\ (E_w(M,K) &= \{x \in M \mid \textit{there is no } y \in M \textit{ such that } x \in y + \text{int } K\} \\ &= M \setminus (M + \text{int } K)). \end{aligned} \tag{9}$$

Definition 2 gives a new representation of $E(M,K)$ that is, the efficient point set of a set M is the difference between two sets. Using this new representation, we prove the properties of $E(M,K)$. Proposition 4 comes from Luc [1], Papageorgiou [3], and Guerraggio et al. [11]. We give an easier proof of this proposition.

Proposition 4. *Assume that M, M_1, and M_2 are nonempty sets and K, K_1, and K_2 are cones in R^n, $\lambda > 0$. Then*

(i) $E(M,K_2) \subseteq E(M,K_1)$, *if* $K_1 \subseteq K_2$,

(ii) $E(\lambda M, K) = \lambda E(M, K)$,

(iii) $E(M_1 + M_2, K) \subseteq E(M_1, K) + E(M_2, K)$,

(iv) $E(M + H, K) = E(M, K)$, *for any* $\{0\} \subseteq H \subseteq K$, K *is a convex cone.*

Proof. For convenience, let $K_{10} = K_1 \setminus \{0\}$, and let $K_{20} = K_2 \setminus \{0\}$. Since $K_1 \subseteq K_2$, we have $M + K_{10} \subseteq M + K_{20}$. And then

$$E(M, K_2) = M \setminus (M + K_{20}) \subseteq M \setminus (M + K_{10}) = E(M, K_1), \tag{10}$$

$$\begin{aligned} E(\lambda M, K) &= (\lambda M) \setminus (\lambda M + K_0) \\ &= (\lambda M) \setminus (\lambda M + \lambda K_0) \\ &= (\lambda M) \setminus \lambda (M + K_0) \\ &= \lambda [M \setminus (M + K_0)] = \lambda [E(M, K)]. \end{aligned} \tag{11}$$

By Theorem 3, it is sufficient to show that

$$(M_1 + M_2) \setminus (M_1 + M_2 + K_0) \subseteq [M_1 \setminus (M_1 + K_0)] + [M_2 \setminus (M_2 + K_0)], \tag{12}$$

for all $x = x_1 + x_2 \in (M_1 + M_2) \setminus (M_1 + M_2 + K_0)$, where $x_1 \in M_1$, $x_2 \in M_2$ and $x_1 + x_2 \notin M_1 + M_2 + K_0$. Then $x_1 \notin M_1 + K_0$, otherwise $x_1 + x_2 \in M_1 + M_2 + K_0$, a contradiction. Similarly $x_2 \notin M_2 + K_0$. Hence

$$x = x_1 + x_2 \in [M_1 \setminus (M_1 + K_0)] + [M_2 \setminus (M_2 + K_0)]. \tag{13}$$

Since $0 \in H \subseteq K$ and K is a convex cone, we have $M \subset M + H$, $H + K_0 = K_0$,

$$\begin{aligned} E(M + H, K) &= (M + H) \setminus (M + H + K_0) \\ &= (M + H) \setminus (M + K_0) \supset M \setminus (M + K_0) \\ &= E(M, K). \end{aligned} \tag{14}$$

On the other hand, For any $x \in E(M + H, K) = (M + H) \setminus (M + K_0)$, $x = m + h$, $m \in M$, $h \in H$. If $h \neq 0$, then $x \in (M + K_0)$, a contradiction. Hence, $x \in M$, $x \notin M + K_0$, that is, $x \in E(M, K)$. □

A spacial case of (ii) in Proposition 4 is that

$$E\left(\overbrace{M + \cdots + M}^{m}, K\right) = E(mM, K) = mE(M, K), \tag{15}$$

if M is convex and K is a cone.

Besides Proposition 4, we state the following properties.

Corollary 5. *Consider the following:*

$$\begin{aligned} &\text{(v) } E(M, K_1 + K_2) \subseteq E(M, K_1) \cap E(M, K_2), \\ &\text{(vi) } E(M, K_1 \cap K_2) \supseteq E(M, K_1) \cup E(M, K_2). \end{aligned} \tag{16}$$

Proposition 6. *Let M, G_1, and G_2 be sets in R^n. Then*

$$M \setminus [M + (G_1 \cup G_2)_0] = [M \setminus (M + G_{10})] \cap [M \setminus (M + G_{20})]. \tag{17}$$

Proof. It is obvious that $(G_1 \cup G_2)_0 = G_{10} \cup G_{20}$ and $M + (G_1 \cup G_2)_0 = M + (G_{10} \cup G_{20}) = (M + G_{10}) \cup (M + G_{20})$. Hence

$$\begin{aligned} M \setminus [M + (G_1 \cup G_2)_0] &= M \setminus [M + (G_{10} \cup G_{20})] \\ &= M \setminus [(M + G_{10}) \cup (M + G_{20})] \\ &= M \cap \overline{(M + G_{10}) \cup (M + G_{20})} \\ &= \left(M \cap \overline{M + G_{10}}\right) \cap \left(M \cap \overline{M + G_{20}}\right) \\ &= [M \setminus (M + G_{10})] \cap [M \setminus (M + G_{20})]. \end{aligned} \tag{18}$$

□

Let $G_1 = K_1$, and let $G_2 = K_2$ in Proposition 6. We have Corollary 7.

Corollary 7. *If K_1 and K_2 are cones, then*

$$E(M, K_1 \cup K_2) = E(M, K_1) \cap E(M, K_2). \tag{19}$$

In the following, we investigate some new properties of the efficient set when M is the union, intersection, sum, or difference of two sets.

Lemma 8. *If K is a convex cone, then*

$$E[M_1 \setminus (M_2 + K), K] \subset E(M_1, K). \tag{20}$$

Proof. If $E[M_1 \setminus (M_2 + K), K] = \emptyset$, the result obviously holds.

Otherwise, for all $x \in E[M_1 \setminus (M_2 + K), K]$, if $x \notin E(M_1, K)$, $\exists y_1 \in M_1$, $k \in K$ such that $x = y_1 + k$, $k \neq 0$, where $y_1 \in M_1 = [M_1 \setminus (M_2 + K)] \cup [M_1 \cap (M_2 + K)]$.

If $y_1 \in M_1 \setminus (M_2 + K)$, it contradicts that $x \in E[M_1 \setminus (M_2 + K), K]$. Hence $y_1 \in M_1 \cap (M_2 + K)$. $\exists y_2 \in M_2$ such that $y_1 \in y_2 + K$, so $x \in y_1 + K \subset y_2 + K \subset M_2 + K$, which still contradicts that $x \in E[M_1 \setminus (M_2 + K), K]$. In consequence, $x \in E(M_1, K)$. □

A more accurate relationship between the two efficient sets in Lemma 8 is described in the following proposition.

Proposition 9. *If K is a convex cone, then*

$$E[M_1 \setminus (M_2 + K), K] = E(M_1, K) \setminus (M_2 + K). \tag{21}$$

Proof. $E[M_1 \setminus (M_2 + K), K] \subset E(M_1, K)$ obtained by Lemma 8.

Note that $(M_1 \setminus M_2 + K) \cap (M_2 + K) = \emptyset$, $E[M_1 \setminus (M_2 + K), K] = E[M_1 \setminus (M_2 + K), K] \cap \overline{(M_2 + K)}$. So $E[M_1 \setminus (M_2 + K), K] = E[M_1 \setminus (M_2 + K), K] \cap \overline{(M_2 + K)} \subset E(M_1, K) \setminus (M_2 + K)$.

Reciprocally, for all $x \in E(M_1, K) \setminus (M_2 + K)$, $x \in E(M_1, K)$, $x \notin M_2 + K$. If $x \notin E[M_1 \setminus (M_2 + K), K]$, $\exists y \in M_1 \setminus (M_2 + K)$ with $x \neq y$ such that $x \in y + K$, in contradiction to $x \in E(M_1, K)$. Therefore,

$$E\left(M_1, K\right) \setminus \left(M_2 + K\right) \subset E\left[M_1 \setminus \left(M_2 + K\right), K\right]. \tag{22}$$

□

Usually $E(M_1 \setminus M_2, K) \neq E(M_1, K) \setminus M_2$, is what Example 10 shows.

Example 10. Consider the following:

$$\begin{gathered} M_1 = \left\{(x, y) \in R^2 \mid 0 \le x \le 1, 0 \le y \le 1\right\}, \\ M_2 = \left\{(x, y) \in R^2 \mid 0 \le x \le 1, 0 \le y \le 1 - x\right\}, \\ M_1 \setminus M_2 = \left\{(x, y) \in R^2 \mid 0 \le x \le 1, 1 - x \le y \le 1\right\}, \\ E\left(M_1, R_2^+\right) = \{(0, 0)\}, \\ E\left(M_1 \setminus M_2, R_2^+\right) \\ = \left\{(x, y) \mid x + y = 1, 0 \le x \le 1, 0 \le y \le 1\right\}. \end{gathered} \tag{23}$$

About the efficient point set of differences between two sets M_1 and M_2, Proposition 11 gives the conclusion, without requiring K to be a convex but a cone.

Proposition 11. *Consider the following:*

$$E\left(M_1 \setminus M_2, K\right) \supseteq \left[E\left(M_1, K\right) \setminus M_2\right] \cup \left[E\left(\overline{M_2}, K\right) \setminus \overline{M_1}\right] \tag{24}$$

Proof. We have

$$\begin{aligned} E\left(M_1 \setminus M_2, K\right) &= \left(M_1 \setminus M_2\right) \setminus \left[\left(M_1 \setminus M_2\right) + K_0\right] \\ &= \left(M_1 \cap \overline{M_2}\right) \cap \overline{\left(M_1 \cap \overline{M_2}\right) + K_0} \\ &\supseteq \left(M_1 \cap \overline{M_2}\right) \cap \overline{\left(M_1 + K_0\right) \cap \left(\overline{M_2} + K_0\right)} \\ &= \left(M_1 \cap \overline{M_2}\right) \cap \left[\overline{M_1 + K_0} \cup \overline{\left(\overline{M_2} + K_0\right)}\right] \\ &= \left[\left(M_1 \cap \overline{M_2}\right) \cap \left(\overline{M_1 + K_0}\right)\right] \\ &\quad \cup \left[\left(M_1 \cap \overline{M_2}\right) \cap \left(\overline{\overline{M_2} + K_0}\right)\right] \\ &= \left[E\left(M_1, K\right) \cap \overline{M_2}\right] \cup \left[E\left(\overline{M_2}, K\right) \cap M_1\right] \\ &= \left[E\left(M_1, K\right) \setminus M_2\right] \cup \left[E\left(\overline{M_2}, K\right) \setminus \overline{M_1}\right]. \end{aligned} \tag{25}$$

□

Remark 12. In the previously mentioned proposition, if $(M_1 \cap \overline{M_2}) + K_0 = (M_1 + K_0) \cap (\overline{M_2} + K_0)$, the equation holds.

Propositions 13 and 15 present the properties of the efficient point sets; when M is the union or intersection of two sets, respectively, K is not required to be a convex.

Proposition 13. *Consider the following:*

$$\begin{aligned} E\left(M_1 \cup M_2, K\right) = {} & \left[E\left(M_1, K\right) \setminus \left(M_2 + K_0\right)\right] \\ & \cup \left[E\left(M_2, K\right) \setminus \left(M_1 + K_0\right)\right]. \end{aligned} \tag{26}$$

Proof. We have

$$\begin{aligned} E\left(M_1 \cup M_2, K\right) &= \left(M_1 \cup M_2\right) \setminus \left[\left(M_1 \cup M_2\right) + K_0\right] \\ &= \left(M_1 \cup M_2\right) \cap \overline{\left(M_1 \cup M_2\right) + K_0} \\ &= \left(M_1 \cup M_2\right) \cap \overline{\left(M_1 + K_0\right) \cup \left(M_2 + K_0\right)} \\ &= \left(M_1 \cup M_2\right) \cap \overline{M_1 + K_0} \cap \overline{M_2 + K_0} \\ &= \left(M_1 \cap \overline{M_1 + K_0} \cap \overline{M_2 + K_0}\right) \\ &\quad \cup \left(M_2 \cap \overline{M_2 + K_0} \cap \overline{M_1 + K_0}\right) \\ &= \left[E\left(M_1, K\right) \setminus \left(M_2 + K_0\right)\right] \\ &\quad \cup \left[E\left(M_2, K\right) \setminus \left(M_1 + K_0\right)\right]. \end{aligned} \tag{27}$$

□

Remark 14. It is obvious that if $E(M_1, K) \cap (M_2 + K_0) = \emptyset$ and $E(M_2, K) \cap (M_1 + K_0) = \emptyset$, then

$$E\left(M_1 \cup M_2, K\right) = E\left(M_1, K\right) \cup E\left(M_2, K\right). \tag{28}$$

Proposition 15. *Consider the following:*

$$E\left(M_1 \cap M_2, K\right) \supseteq \left[E\left(M_1, K\right) \cap M_2\right] \cup \left[E\left(M_2, K\right) \cap M_1\right]. \tag{29}$$

Proof. Notice that $(M_1 \cap M_2) + K_0 \subseteq (M_1 + K_0) \cap (M_2 \cap K_0)$.

$$\begin{aligned} E\left(M_1 \cap M_2, K\right) &= \left(M_1 \cap M_2\right) \setminus \left[\left(M_1 \cap M_2\right) + K_0\right] \\ &\supseteq \left(M_1 \cap M_2\right) \setminus \left[\left(M_1 + K_0\right) \cap \left(M_2 + K_0\right)\right] \\ &= \left(M_1 \cap M_2\right) \cap \overline{\left(M_1 + K_0\right) \cap \left(M_2 + K_0\right)} \\ &= \left(M_1 \cap M_2\right) \cap \left[\overline{M_1 + K_0} \cup \overline{M_2 + K_0}\right] \\ &= \left[\left(M_1 \cap M_2\right) \cap \overline{M_1 + K_0}\right] \\ &\quad \cup \left[\left(M_1 \cap M_2\right) \cap \overline{M_2 + K_0}\right] \\ &= \left[E\left(M_1, K\right) \cap M_2\right] \cup \left[E\left(M_2, K\right) \cap M_1\right]. \end{aligned} \tag{30}$$

□

Remark 16. (i) Usually the equality does not hold in Proposition 15. Let $M_1 = \{(x, y) \mid 0 \le x \le y \le 1\}$, $M_2 = \{(x, y) \mid 0 \le x - 1/2 \le y \le 1\}$. Then

$$\begin{gathered} E\left(M_1 \cap M_2, K\right) = \{(1/2, 1/2)\}, \\ \left[E\left(M_1, K\right) \cap M_2\right] \cup \left[E\left(M_2, K\right) \cap M_1\right] = \emptyset, \end{gathered} \tag{31}$$

(ii) for Proposition 15, if $(M_1 \cap M_2) + K_0 = (M_1 + K_0) \cap (M_2 + K_0)$, then

$$E(M_1 \cap M_2, K) = [E(M_1, K) \cap M_2] \cup [E(M_2, K) \cap M_1], \tag{32}$$

(iii) since $M_1 \setminus M_2 = M_1 \cap \overline{M_2}$, Proposition 11 can be obtained by Proposition 15.

3. Efficiency of Four Most Representative DEA Models

For each of DEA models, Charnes et al. and Wei et al. (e.g., see [5, 7, 9]) establish an associative multiobjective linear programming model and prove that a DMU_{j_0} is DEA efficient if and only if (x_{j_0}, y_{j_0}) (associating with the DMU_{j_0}) is a Pareto efficient solution of the multiobjective linear programming problem (for the details about DEA efficient of DMUs in DEA model, see [4, 5] and the references therein). This beautiful conclusion provides the multiobjective linear programming as an efficient tool to solve DEA problems. So the key point now is to find all efficient solutions of the multiobjective linear programming problems. This section offers a simple way to do this.

By the results of Section 2, in the following, we investigate the multiobjective linear programming problems corresponding to the four DEA models and develop the structures of the efficient point sets and the efficient solution sets of these programmings. Based on these, the relationships between the DEA efficiency of the four DEA models are obtained.

Denote $x_j = (x_{1j}, \dots, x_{mj}) > 0 (x_j \in R^m)$ to be the input vector for the jth decision making unit (DMU_j), and $y_j = (y_{1j}, \dots, y_{sj}) > 0$ $(y_j \in R^s)$, to be the output vector for the jth decision making unit, for $j = 1, \dots, n$. For convenience the notations $x_0 = x_{j_0}$, $y_0 = y_{j_0}$, for $1 \le j_0 \le n$ are given. Let $C = R_+^m \subset R^m$, and let $D = R_+^s \subset R^s$. The orderings in R^m and R^s are defined by C and D, respectively. $K = \{(c, d) \mid c \in C, d \in D\}$, $K_1 = \{(c, -d) c \in C, d \in D\}$, $K_0 = K \setminus \{0\}$, and $K_{10} = K_1 \setminus \{0$. The production possibility set T, $T \subset \{(x, y) \mid x \in R^m, \ y \in R^s, x \ge 0, y \ge 0\}$, is based on postulate sets which are presented with a brief explanation (see [4, 5]). The four most representative models are, briefly, C^2R, BC^2, FG, and ST, which correspond to different production possibility sets T_{C^2R}, T_{BC^2}, T_{FG}, and T_{ST}, respectively, [4–10]

$$\begin{aligned} T_{C^2R} = & \left\{ (x, y) \mid \sum_{j=1}^{n} x_j \lambda_j \le x, \right. \\ & \left. \sum_{j=1}^{n} y_j \lambda_j \ge y \ge 0, \ \lambda_j \ge 0, \ j = 1, \dots, n \right\}, \\ T_{BC^2} = & \left\{ (x, y) \mid \sum_{j=1}^{n} x_j \lambda_j \le x, \ \sum_{j=1}^{n} y_j \lambda_j \ge y \ge 0, \right. \\ & \left. \sum_{j=1}^{n} \lambda_j = 1, \ \lambda_j \ge 0, \ j = 1, \dots, n \right\}, \\ T_{FG} = & \left\{ (x, y) \mid \sum_{j=1}^{n} x_j \lambda_j \le x, \ \sum_{j=1}^{n} y_j \lambda_j \ge y \ge 0, \right. \\ & \left. \sum_{j=1}^{n} \lambda_j \le 1, \ \lambda_j \ge 0, \ j = 1, \dots, n \right\}, \\ T_{ST} = & \left\{ (x, y) \mid \sum_{j=1}^{n} x_j \lambda_j \le x, \ \sum_{j=1}^{n} y_j \lambda_j \ge y \ge 0, \right. \\ & \left. \sum_{j=1}^{n} \lambda_j \ge 1, \ \lambda_j \ge 0, \ j = 1, \dots, n \right\}. \end{aligned} \tag{33}$$

Obviously, the following equalities and inclusions hold:

$$\begin{aligned} &\text{(a)} \ T_{C^2R} = T_{FG} \cup T_{ST}, \qquad \text{(b)} \ T_{BC^2} = T_{FG} \cap T_{ST}, \\ &\text{(c)} \ T_{C^2R} \supseteq T_{FG} \supseteq T_{BC^2}, \qquad \text{(d)} \ T_{C^2R} \supseteq T_{ST} \supseteq T_{BC^2}. \end{aligned} \tag{34}$$

For distinguishing the DEA efficiency of DMUs in DEA models, according to the equivalence of DEA efficiency in DEA models and Pareto efficiency in multiobjective linear programming obtained by Charnes et al. and Wei et al., we introduce the multiobjective linear programming problem corresponding to DEA models in the following. For the details about DEA models, see [7, 8].

Consider multiobjective linear programming problem:

$$\begin{aligned} & V\text{-}\min (x, -y), \\ & \text{s.t.} \ (x, y) \in T. \end{aligned} \tag{VP}$$

For $T = T_{C^2R}$, T_{BC^2}, T_{FG}, or T_{ST}, the four multiobjective linear programming problems VP_{C^2R}, VP_{BC^2}, VP_{FG}, and VP_{ST} correspond to the four DEA models respectively.

Definition 17. (x_0, y_0) is called an Pareto efficient (weak efficient) solution of (VP), if there is no $(x, y) \in T$, $(x, y) \ne (x_0, y_0)$ such that

$$\begin{aligned} & (x, -y) \le (x_0, -y_0), \\ & ((x, -y) < (x_0, -y_0)). \end{aligned} \tag{35}$$

Since K defines the ordering in $R^{m+s} (\supset T)$, the previous inequality can be written as

$$\begin{aligned} & (x_0, -y_0) \in (x, -y) + K_0, \\ & ((x_0, -y_0) \in (x, -y) + \operatorname{int} K). \end{aligned} \tag{36}$$

When K is a convex pointed cone, all conclusions in Section 2 hold for multiobjective programming problem (VP). Theorem 18 provides the equivalency of DEA efficiency of DMUs and the Pareto efficiency of the points corresponding to the DMUs. Consequently, the structures of the sets of efficient solutions to DEA models are obtained by solving the multiobjective programmings. Then the relationships between the DEA efficiencies of DMUs in DEA models are deduced.

Theorem 18 (see [5]). *A* DMU_{j_0} *is DEA efficient if and only if* (x_0, y_0) *is a Pareto efficient solution of* (VP).

For convenience, we denote by DMU_0 DMU_{j_0}. In the following, we investigate the structures of the efficient point sets and the efficient solution sets of the four vector optimization problems. Let

$$\begin{aligned} M_0 &= \left\{\left(x_j, -y_j\right) \mid j = 1, \ldots, n\right\}, \\ T_0 &= \left\{\left(x_j, y_j\right) \mid j = 1, \ldots, n\right\}, \\ M_{C^2R} &= \left\{(x, -y) \mid (x, y) \in T_{C^2R}\right\}, \\ M_{BC^2} &= \left\{(x, -y) \mid (x, y) \in T_{BC^2}\right\}, \\ M_{FG} &= \left\{(x, -y) \mid (x, y) \in T_{FG}\right\}, \\ M_{ST} &= \left\{(x, -y) \mid (x, y) \in T_{ST}\right\}. \end{aligned} \tag{37}$$

The relations (a), (b), (c), and (d) imply:

$$\begin{aligned} &(\bar{a})\ \ M_{C^2R} = M_{FG} \cup M_{ST}, \\ &(\bar{b})\ \ M_{BC^2} = M_{FG} \cap M_{ST}, \\ &(\bar{c})\ \ M_{C^2R} \supseteq M_{FG} \supseteq M_{BC^2}, \\ &(\bar{d})\ \ M_{C^2R} \supseteq M_{ST} \supseteq M_{BC^2}. \end{aligned} \tag{38}$$

As an example, consider the multiobjective linear programming problem (VP_{C^2R}). Let E_{C^2R} denote the efficient point set, and let S_{C^2R} denote the efficient solution set. Note that

$$(x, -y) \leq (x_0, -y_0) \Longleftrightarrow (x_0, -y_0) \in (x, -y) + K. \tag{39}$$

By Definitions 2 and 17,

$$E_{C^2R} = M_0 \setminus \left(M_{C^2R} + K_0\right). \tag{40}$$

Since

$$\left(x_0, -y_0\right) \in (x, -y) + K_0 \Longleftrightarrow \left(x_0, y_0\right) \in (x, y) + K_{10}, \tag{41}$$

we have

$$S_{C^2R} = T_0 \setminus \left(T_{C^2R} + K_{10}\right). \tag{42}$$

Similar argument is applied to other three multiobjective linear programming problems. Therefore the following two theorems hold.

Theorem 19. *Consider the following:*

$$\begin{aligned} &(\bar{\bar{a}})\ \ E_{C^2R} = M_0 \setminus \left(M_{C^2R} + K_0\right), \\ &(\bar{\bar{b}})\ \ E_{BC^2} = M_0 \setminus \left(M_{BC^2} + K_0\right), \\ &(\bar{\bar{c}})\ \ E_{FG} = M_0 \setminus \left(M_{FG} + K_0\right), \\ &(\bar{\bar{d}})\ \ E_{ST} = M_0 \setminus \left(M_{ST} + K_0\right). \end{aligned} \tag{43}$$

Theorem 20. *Consider the following:*

$$\begin{aligned} &(\bar{\bar{\bar{a}}})\ \ S_{C^2R} = T_0 \setminus \left(T_{C^2R} + K_{10}\right), \\ &(\bar{\bar{\bar{b}}})\ \ S_{BC^2} = T_0 \setminus \left(T_{BC^2} + K_{10}\right), \\ &(\bar{\bar{\bar{c}}})\ \ S_{FG} = T_0 \setminus \left(T_{FG} + K_{10}\right), \\ &(\bar{\bar{\bar{d}}})\ \ S_{ST} = T_0 \setminus \left(T_{ST} + K_{10}\right). \end{aligned} \tag{44}$$

Inclusions (c), (d) and Theorem 20 imply Theorem 21.

Theorem 21. *Consider the following:*

$$\begin{aligned} &\text{(i)}\ \ S_{C^2R} \subseteq S_{FG} \subseteq S_{BC^2}, \\ &\text{(ii)}\ \ S_{C^2R} \subseteq S_{ST} \subseteq S_{BC^2}. \end{aligned} \tag{45}$$

Theorem 21 shows the relationships between the solutions of the four vector optimization problems. Therefore, by Theorem 18, the relationships between DEA efficiencies of DMUs in DEA models are as the following:

(i) if DMU_0 is C^2R DEA efficient, then it is FG DEA efficient. If DMU_0 is FG DEA efficient, then it is B^2C DEA efficient;

(ii) if DMU_0 is C^2R DEA efficient, then it is ST DEA efficient. If DMU_0 is ST DEA efficient, then it is B^2C DEA efficient.

By Proposition 13, along with the equality $(\bar{a})$ and $(\bar{\bar{a}})$, we obtain following Theorem 22.

Theorem 22. *Consider the following:*

$$\begin{aligned} E_{C^2R} &= \left[M_0 \setminus \left(M_{FG} + K_0\right)\right] \setminus \left(M_{ST} + K_0\right) \\ &= \left[M_0 \setminus \left(M_{ST} + K_0\right)\right] \setminus \left(M_{FG} + K_0\right) \\ &= E_{FG} \setminus \left(M_{ST} + K_0\right) = E_{ST} \setminus \left(M_{FG} + K_0\right). \end{aligned} \tag{46}$$

By Theorem 22, we conclude that if $E_{C^2R} \neq \phi$, then $E_{FG} \neq \phi$ and $E_{ST} \neq \phi$; by $(\bar{\bar{\bar{a}}})$, $(\bar{\bar{\bar{c}}})$, and $(\bar{\bar{\bar{d}}})$, we declaim that if DMU_0 is C^2R DEA efficient, then it is also FG DEA efficient and ST DEA efficient.

Proposition 15 and $(\bar{b})$ infer Theorem 23.

Theorem 23. *Consider the following:*

$$E_{BC^2} \supseteq \left(E_{FG} \cap M_{ST}\right) \cup \left(E_{ST} \cap M_{FG}\right). \tag{47}$$

Consider that $E_{FG} = M_0 \setminus (M_{FG} + K_0)$, and $M_0 \subset M_{ST}$, if $E_{FG} \neq \phi$, then $E_{FG} \cap M_{ST} \neq \phi$. Similarly, $E_{ST} \cap M_{FG} \neq \phi$. Therefore, the following consequences are obtained:

(i) if DMU_0 is ST DEA efficient, then it is also BC^2 DEA efficient;

(ii) if DMU_0 is FG DEA efficient, then it is also BC^2 DEA efficient.

4. Conclusion

In this paper, Definition 2 presents a new representation to $E(M, K)$. Then some new properties of $E(M, K)$ are deduced by the new representation (Propositions 6 to 15 and their corollaries). Most of the properties are proved by mean of the binary operations of sets, which is a new method in multiobjective optimization. These conclusions are used to deal with the multiobjective linear programming problems corresponding to the four most representative DEA models. We investigate the structures of the efficient solution set of the four DEA models (Theorems 19 and 20) and deduce the relationships between DEA efficient solution sets of the four DEA models (Theorems 21 and 22). For each of the four DEA models, by Theorem 20, $S = T_0 \setminus (T + K_{10})$; that is, all the DEA efficient DMUs are obtained by the binary operations of sets. Therefore, this is a simple way to get all DEA efficient DMUs theoretically. By using the new representation, it may be able to discuss the effects upon DEA efficiency of DMUs when the number of the DMUs changes in DEA models.

Acknowledgments

This research was supported in part by the Zhejiang Natural Science Foundation of China Grant Y6110054, the Shanxi Scholarship Council of China Research Grant 2010087.

References

[1] D. T. Luc, *Theory of Vector Optimization*, vol. 319, Springer, Berlin, Germany, 1989.

[2] J. Jahn, *Vector Optimization, Theory, Applications, and Extensions*, Springer, Berlin, Germany, 2004.

[3] N. S. Papageorgiou, "Pareto efficiency in locally convex spaces. I," *Numerical Functional Analysis and Optimization*, vol. 8, no. 1-2, pp. 83–116, 1985.

[4] W. W. Cooper, L. M. Seiford, and K. Tone, *Data Envelopment Analysis, A Comprehensive Text With Models, Applications, References and DEA-Solver Software*, Springer, 2007.

[5] Q. L. Wei, *Data Envelopment Analysis*, Science, Beijing, Germany, 2004.

[6] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978.

[7] A. Charnes, W. W. Cooper, Q. L. Wei, and Z. M. Huang, "Cone ratio data envelopment analysis and multi-objective programming," *International Journal of Systems Science*, vol. 20, no. 7, pp. 1099–1118, 1989.

[8] Q. L. Wei and G. Yu, "Analyzing properties of K-cones in the generalized data envelopment analysis model," *Journal of Econometrics*, vol. 80, no. 1, pp. 63–84, 1997.

[9] Q. L. Wei, "Equivalence between weak Pareto and weak DEA efficiency in input-oriented and output-oriented DEA model," *Systems Engineering-Theory & Practice*, vol. 22, pp. 72–80, 2002.

[10] Q. L. Wei, H. Yan, and G. Hao, "Characteristics and structures of weak efficient surfaces of production possibility sets," *Journal of Mathematical Analysis and Applications*, vol. 327, no. 2, pp. 1055–1074, 2007.

[11] A. Guerraggio, E. Molho, and A. Zaffaroni, "On the notion of proper efficiency in vector optimization," *Journal of Optimization Theory and Applications*, vol. 82, no. 1, pp. 1–21, 1994.