Mathematical Problems in Engineering

Theory, Methods, and Applications

Editor-in-Chief: Jose Manoel Balthazar

Special Issue
Short Range Phenomena: Modeling, Computational Aspects
and Applications

Guest Editors: Carlo Cattani, Ming Li, and Cristian Toma

Hindawi Publishing Corporation
http://www.hindawi.com

Volume 2008

Short Range Phenomena: Modeling, Computational Aspects, and Applications

Mathematical Problems in Engineering

Short Range Phenomena: Modeling, Computational Aspects, and Applications

Guest Editors: Carlo Cattani, Ming Li, and Cristian Toma

Editor-in-Chief

Associate Editors

Contents

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 761081, 2 pages
doi:10.1155/2008/761081

Editorial

Short Range Phenomena: Modeling, Computational Aspects, and Applications

Carlo Cattani,[1] Ming Li,[2] and Cristian Toma[3]

[1] *Department of Pharmaceutical Sciences (DiFarma), University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy*
[2] *Department of Electronic Science and Technology, School of Information Science and Technology, East China Normal University, Shanghai 200062, China*
[3] *Faculty of Applied Sciences, Politehnica University, 060042 Bucharest, Romania*

Correspondence should be addressed to Carlo Cattani, ccattani@unisa.it

Received 16 September 2008; Accepted 16 September 2008

In the recent years, the mathematical formalism of impulsive systems (based on impulsive differential equations) has tried to join together the rigorous aspects from continuous systems formalism and the wide range of applications of discrete systems formalism. They were introduced to handle many evolution processes which are subject to singular short-term perturbations. Abrupt changes must be approached with logical, mathematical, and technical aspects dealing with the final evolution of such impulsive sources, whose effects are entirely transferred to the new state of the systems. Modern aspects in physics (quantum theory) and mathematics (wavelets, fractal theory) should be expedient in modelling short range phenomena, and describing dynamics of perturbations and transitions in natural systems (advanced materials science) and advanced systems (optic, electronic, and quantum devices).

The aim of this special issue is to present recent advances of theoretical, computational, and practical aspects for modeling short range phenomena in order to reveal new fundamental aspects in science and engineering. Using mathematical tools of wavelets analysis, fractal theory, and applied mathematics (signal processing, numerical simulations, control theory) adapted for short range phenomena, significant results were obtained in the research fields of structure analysis and image recognition, wavelets analysis of localised space-time phenomena, dynamical and computational aspects of pulse measurement, sequences of pulses and time series, and mathematical and physical aspects of pulse generation.

This special issue involves 15 original papers, selected by the editors so as to present the most significant results in the previously mentioned topics. These papers are organised as follows.

(a) Three papers on structure analysis and image recognition: "Incremental nonnegative matrix factorization for face recognition" by Wen-Sheng Chen et al., "Direct neighborhood discriminant analysis for face recognition" by Miao Cheng et al., and "Intelligent control of the complex technology process based on adaptive pattern clustering and feature map" by Cheng Wushan.

(b) Three papers on wavelets analysis of localised space-time phenomena: "Shannon wavelets theory" by Carlo Cattani "Combined Preorder and Postorder Traversal Algorithm for the Analysis of Singular Systems by Haar Wavelets" by Beom-Soo Kim et al., and "On the discrete harmonic wavelet transform" by Carlo Cattani and Aleksey Kudreyko.

(c) Three papers on sequences of pulses and time series: "Resolution of first and second order linear differential equations with periodic inputs by a computer algebra system" by Matilde Legua et al., "Detection of variations of local irregularity of traffic under DDOS flood attack" by Ming Li and Wei Zhao, and "Tool wear detection based on Duffing-Holmes oscillator" by Wanqing Song et al.

(d) Three papers on dynamical and computational aspects of pulse measurement: "Venturi wet gas flow modeling based on homogenous and separated flow theory" by Fang Lide et al., "Detection of short step pulses using practical test-functions and resonance aspects" by Alexandru Toma and Cristian Morarescu, and "On nonperturbative techniques for thermal radiation effect on natural convection past a vertical plate embedded in a saturated porous medium" by Oluwole Makinde and Raseelo J. Moitsheki.

(e) Two papers on mathematical and physical aspects of pulse generation: "Relativistic short range phenomena and space-time aspects of pulse measurements" by Ezzat Bakhoum and Cristian Toma and "Vanishing waves on closed intervals and propagating short-range phenomena" by Toma Ghiocel and Flavia Doboga.

(f) One paper on applications of short range (localised) phenomena analysis in biology: "Solving ratio-dependent predator-prey system with constant effort harvesting using homotopy perturbation method" by Abdoul Reza Ghotbi et al.

Carlo Cattani
Ming Li
Cristian Toma

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 410674, 17 pages
doi:10.1155/2008/410674

Research Article

Incremental Nonnegative Matrix Factorization for Face Recognition

Wen-Sheng Chen,[1] Binbin Pan,[1] Bin Fang,[2] Ming Li,[3] and Jianliang Tang[1]

[1] *College of Mathematics and Computational Science, Shenzhen University, Shenzhen 518060, China*
[2] *College of Computer Science, Chongqing University, Chongqing 400044, China*
[3] *School of Information Science & Technology, East China Normal University, Shanghai 200241, China*

Correspondence should be addressed to Wen-Sheng Chen, chenws@szu.edu.cn

Received 25 May 2008; Accepted 5 June 2008

Recommended by Cristian Toma

Nonnegative matrix factorization (NMF) is a promising approach for local feature extraction in face recognition tasks. However, there are two major drawbacks in almost all existing NMF-based methods. One shortcoming is that the computational cost is expensive for large matrix decomposition. The other is that it must conduct repetitive learning, when the training samples or classes are updated. To overcome these two limitations, this paper proposes a novel incremental nonnegative matrix factorization (INMF) for face representation and recognition. The proposed INMF approach is based on a novel constraint criterion and our previous block strategy. It thus has some good properties, such as low computational complexity, sparse coefficient matrix. Also, the coefficient column vectors between different classes are orthogonal. In particular, it can be applied to incremental learning. Two face databases, namely FERET and CMU PIE face databases, are selected for evaluation. Compared with PCA and some state-of-the-art NMF-based methods, our INMF approach gives the best performance.

1. Introduction

Face recognition has been one of the most challenging problems in computer science and information technology since 1990 [1, 2]. The approaches of face recognition can be mainly categorized into two groups, namely geometric feature-based and appearance-based [3]. The geometric features are based on the short range phenomena of face images such as eyes, eyebrows, nose, and mouth. The facial local features are learnt to form a face geometric feature vector for face recognition. The appearance-based approach relies on the global facial features, which generate an entire facial feature vector for face classification. Nonnegative matrix factorization (NMF) [4, 5] belongs to geometric feature-based category, while principle component analysis (PCA) [6] is based on the whole facial features. Both NMF and PCA are unsupervised learning methods for face recognition. The basic ideas of these two approaches are to find the basis images using different criterions. All face images can be reconstructed by the basis images. The basis images of PCA are called eigenfaces, which are the eigenvectors corresponding to large eigenvalues of total scatter matrix. NMF aims to perform nonnegative matrix decomposition on the training image matrix V such that $V \approx WH$, where W and H are the basis image matrix and the coefficient matrix, respectively. The local image features are learnt and contained in W as column vectors. Follow the success of applying NMF in learning the parts of objects [4], many researchers have conducted in-depth investigation on NMF and different NMF-based approaches have been developed [7–19]. Li et al. proposed a local NMF method [7] by adding some spatial constraints. Wild et al. [8] utilized spherical K-means clustering to produce a structured initialization for NMF. Buciu and Pitas [9] presented a DNMF method for learning facial expressions in a supervised manner. However, DNMF does not guarantee convergence to a stationary limit point. Kotsia et al. [15] thus presented a modified DNMF method using projected gradients. Some similar supervised methods incorporated into NMF were developed to enhance the classification power of NMF [11–13, 19]. Hoyer [10] added sparseness constraints to NMF to find solutions with desired degrees of sparseness. Lin [16, 17] modified traditional NMF updates using projected gradient method and discussed their convergences. Recently, Zhang et al. [18] proposed a topology structure preservation constraint in NMF to improve the NMF performance.

However, to the best of our knowledge, almost all existing NMF-based approaches encounter two major problems, namely time-consuming problem and incremental learning problem. In most cases, the training image matrix V is very large and it leads to expensive computational cost for NMF-based schemes. Also, when the training samples or classes are updated, NMF must implement repetitive learning. These drawbacks greatly restrict the practical applications of NMF-based methods to face recognition. To avoid the above two problems, this paper, motivated by our previous work on incremental learning [19], proposes a supervised incremental NMF (INMF) approach under a novel constraint NMF criterion, which aims to cluster within class samples tightly and augment the between-class distance simultaneously. Our incremental strategy utilizes the supervised local features, which are considered as the short-range phenomena of face images, for face classifications. Two public available face databases, namely FERET and CMU PIE face databases, are selected for evaluation. Experimental results show that our INMF method outperforms PCA [6], NMF [4], and BNMF [19] approaches in both nonincremental learning and incremental learning of face recognition.

The rest of this paper is organized as follows: Section 2 briefly reviews the related works. Theoretical analysis and INMF algorithm design are given in Section 3. Experimental results are reported in Section 4. Finally, Section 5 draws the conclusions.

2. Related work

This section briefly introduces PCA [6], NMF [4], and BNMF [19] methods. Details are as follows.

2.1. PCA

Principal component analysis (PCA), also called eigenface method, is a popular statistic appearance-based linear method for dimensionality reduction in face recognition. The theory used in PCA is based on Karhunen-Loeve transform. It performs the eigenvalue decomposition on the total scatter matrix S_t and then selects the large principal components (eigenfaces) to account for most distributions. All face images can be expressed by the linear combinations of these basis images (eigenfaces). However, PCA is not able to exploit all of the feature classification information and how to choose the principal component elements is still a problem. Therefore, PCA cannot give satisfactory performance in pattern recognition tasks.

2.2. NMF

NMF aims to find nonnegative matrices W and H such that

$$V_{m\times n} \overset{\text{NMF}}{\approx} W_{m\times r} H_{r\times n}, \tag{2.1}$$

where matrix V is also a nonnegative matrix generated by total n training images. Each column of W is called basis image, while H is the coefficient matrix. The basis number r is usually chosen less than n for dimensionality reduction. The divergence between V and WH is defined as

$$F = \sum_{ij}\left(V_{ij}\log\frac{V_{ij}}{(WH)_{ij}} - V_{ij} + (WH)_{ij}\right). \tag{2.2}$$

NMF (2.1) is equivalent to the following optimization problem:

$$\min_{W,H} F, \quad \text{s.t. } W \geq 0,\ H \geq 0, \qquad \sum_i W_{ik} = 1, \quad \forall\ k. \tag{2.3}$$

The minimization problem (2.3) can be solved using the following iterative formulae, which converge to a local minimum:

$$W_{ij} \longleftarrow W_{ij}\sum_k \frac{V_{ik}}{(WH)_{ik}} H_{jk}, \qquad W_{ij} \longleftarrow \frac{W_{ij}}{\sum_k W_{kj}}, \qquad H_{ij} \longleftarrow H_{ij}\sum_k W_{ki}\frac{V_{kj}}{(WH)_{kj}}. \tag{2.4}$$

2.3. BNMF

The basic idea of BNMF is to perform NMF on c small matrices $V^{(i)} \in \mathbb{R}^{m\times n_0}$ $(i = 1, 2, \ldots, c)$, namely

$$\left(V^{(i)}\right)_{m\times n_0} \overset{\text{NMF}}{\approx} \left(W^{(i)}\right)_{m\times r_0}\left(H^{(i)}\right)_{r_0\times n_0}, \quad i = 1, 2, \ldots, c, \tag{2.5}$$

where $V^{(i)}$ contains n_0 training images of the ith class, and c is the number of classes. BNMF is yielded from (2.5) as follows:

$$V_{m\times n} \overset{\text{BNMF}}{\approx} W_{m\times r} H_{r\times n}, \tag{2.6}$$

where $r = cr_0$, $V_{m\times n} = [\, V^{(1)}\ V^{(2)}\ \cdots\ V^{(c)}\,]$, $W_{m\times r} = [\, W^{(1)}\ W^{(2)}\ \cdots\ W^{(c)}\,]$, $H_{r\times n} = \operatorname{diag}(H^{(1)}, H^{(2)}, \ldots, H^{(c)})$, and $n(= cn_0)$ is the total number of training images.

3. Proposed INMF

To overcome the drawbacks of existing NMF-based methods, this section proposes a novel incremental NMF (INMF) approach, which is based on a new constraint NMF criterion and our previous block technique [19]. Details are discussed below.

3.1. Constraint NMF criterion

The objective of our INMF is to impose supervised class information on NMF such that between-class distances increase, while the within-class distances simultaneously decrease. To this end, we define the within-class scatter matrix $S_w^{(i)}$ of the ith coefficient matrix $H^{(i)} \in \mathbb{R}^{r_0 \times n_0}$ as

$$S_w^{(i)} = \frac{1}{n_0} \sum_{j=1}^{n_0} \left(H_j^{(i)} - U^{(i)}\right)\left(H_j^{(i)} - U^{(i)}\right)^T, \tag{3.1}$$

where $U^{(i)} = (1/n_0)\sum_{j=1}^{n_0} H_j^{(i)}$ is the mean column vector of the ith class. The within-class samples of the kth class will cluster tightly as $\operatorname{tr}(S_w^{(k)})$ becomes small.

Assume $\widetilde{U}^{(i)}$ is an enlarging vector of $U^{(i)}$, that is, $\widetilde{U}^{(i)} = (1+t)U^{(i)}$ with $t > 0$. Then we have

$$\left\|U^{(i)} - U^{(j)}\right\| < (1+t)\left\|U^{(i)} - U^{(j)}\right\| = \left\|\widetilde{U}^{(i)} - \widetilde{U}^{(j)}\right\|. \tag{3.2}$$

Inequality (3.2) implies that between-class distances are increased as the mean vectors of classes in H are enlarged.

Based on above analysis, we define a constraint divergence criterion function for the kth class as follows:

$$F^{(k)} = \sum_{ij} \left(V_{ij}^{(k)} \log \frac{V_{ij}^{(k)}}{(WH)_{ij}^{(k)}} - V_{ij}^{(k)} + (WH)_{ij}^{(k)} \right) + \alpha \operatorname{tr}\left(S_w^{(k)}\right) - \beta \left\|U^{(k)}\right\|_2^2, \tag{3.3}$$

where parameters $\alpha, \beta > 0$ and $k = 1, 2, \ldots, c$.

Our entire INMF criterion function is then designed as below:

$$F = \sum_{k=1}^{c} F^{(k)} = \sum_{ijk} \left(V_{ij}^{(k)} \log \frac{V_{ij}^{(k)}}{(WH)_{ij}^{(k)}} - V_{ij}^{(k)} + (WH)_{ij}^{(k)} \right) + \sum_{k} \left(\alpha \operatorname{tr}\left(S_w^{(k)}\right) - \beta \left\|U^{(k)}\right\|_2^2 \right). \tag{3.4}$$

Based on criterion (3.4), the following constraint NMF (CNMF) update rules (3.5)–(3.7) will be derived in the next subsection. We can show that the iterative formulae (3.5)–(3.7) converge to a local minimum as well:

$$W_{ij}^{(k)} \longleftarrow W_{ij}^{(k)} \sum_{l} \frac{V_{il}^{(k)}}{(WH)_{il}^{(k)}} H_{jl}^{(k)}, \tag{3.5}$$

$$W_{ij}^{(k)} \longleftarrow \frac{W_{ij}^{(k)}}{\sum_l W_{lj}^{(k)}}, \tag{3.6}$$

$$H_{ij}^{(k)} \longleftarrow \frac{-b + \sqrt{b^2 - 4ad}}{2a}, \tag{3.7}$$

where $a = 2\alpha - \beta/n_k^2$, $b = -(2\alpha + \beta/n_k)U_i^{(k)} + 1$, $d = -H_{ij}^{(k)} \sum_l V_{lj}^{(k)} W_{li}^{(k)} / (W^{(k)}H^{(k)})_{lj}$, and $U_i^{(k)}$ is the ith entry of vector $U^{(k)}$, $k = 1, 2, \ldots, c$.

So, our entire INMF is performed as follows:

$$\begin{bmatrix} V^{(1)} & V^{(2)} & \cdots & V^{(c)} \end{bmatrix} \overset{\text{INMF}}{\approx} \begin{bmatrix} W^{(1)} & W^{(2)} & \cdots & W^{(c)} \end{bmatrix} \begin{bmatrix} H^{(1)} & & & \\ & H^{(2)} & & \\ & & \ddots & \\ & & & H^{(c)} \end{bmatrix}, \tag{3.8}$$

where

$$\left(V^{(i)}\right)_{m\times n_0} \overset{\text{CNMF}}{\approx} \left(W^{(i)}\right)_{m\times r_0} \left(H^{(i)}\right)_{r_0\times n_0}, \quad i = 1, 2, \ldots, c. \tag{3.9}$$

3.2. Convergence of proposed constraint NMF

This subsection reports how to derive the iterative formulae (3.5)–(3.7) and discusses their convergences under constraint NMF criterion (3.3).

Definition 3.1 (see [5]). $J(Q, \widetilde{Q})$ is called an auxiliary function for $E(Q)$, if $J(Q, \widetilde{Q})$ satisfies

$$J(Q, \widetilde{Q}) \geq E(Q), \qquad J(Q, Q) = E(Q), \tag{3.10}$$

where Q, $\widetilde{Q}$ are matrices with the same size.

Lemma 3.2 (see [5]). *If $J(Q, \widetilde{Q})$ is an auxiliary function for $E(Q)$, then $E(Q)$ is a nonincreasing function under the update rule*

$$Q^{i+1} = \arg\min_{Q} J(Q, Q^i). \tag{3.11}$$

To obtain iterative rule (3.7) *and prove its convergence, one first constructs an auxiliary function for F with fixed W.*

Theorem 3.3. *If $F^{(k)}(H^{(k)})$ is the value of criterion function* (3.3) *with fixed $W^{(k)}$, then $G^{(k)}$ $(H^{(k)}, \widetilde{H}^{(k)})$ is an auxiliary function for $F^{(k)}(H^{(k)})$, where*

$$\begin{aligned} G^{(k)}\left(H^{(k)}, \widetilde{H}^{(k)}\right) = {} & \sum_{ij} \left(V_{ij}^{(k)} \log V_{ij}^{(k)} - V_{ij}^{(k)} + \left(W^{(k)}H^{(k)}\right)_{ij}\right) \\ & - \sum_{ijl} V_{ij}^{(k)} \frac{W_{il}^{(k)} \widetilde{H}_{lj}^{(k)}}{\sum_l W_{il}^{(k)} \widetilde{H}_{lj}^{(k)}} \left(\log\left(W_{il}^{(k)} H_{lj}^{(k)}\right) - \log \frac{W_{il}^{(k)} \widetilde{H}_{lj}^{(k)}}{\sum_l W_{il}^{(k)} \widetilde{H}_{lj}^{(k)}} \right) \\ & + \alpha\, tr\left(S_w^{(k)}\right) - \beta \left\|U^{(k)}\right\|_2^2. \end{aligned} \tag{3.12}$$

Proof. It can be directly verified that $G^{(k)}(H^{(k)}, H^{(k)}) = F^{(k)}(H^{(k)})$. So we just need show the inequality $G^{(k)}(H^{(k)}, \widetilde{H}^{(k)}) \geq F^{(k)}(H^{(k)})$. To this end, we will use the convex function $y = \log x$. For all i, j, and $\sum_l \sigma_{ijl} = 1$, it holds that

$$-\log\left(\sum_l W_{il}^{(k)} H_{lj}^{(k)}\right) \leq -\sum_l \sigma_{ijl} \log \frac{W_{il}^{(k)} H_{lj}^{(k)}}{\sigma_{ijl}}. \tag{3.13}$$

Substituting $\sigma_{ijl} = W_{il}^{(k)}\widetilde{H}_{lj}^{(k)} / \sum_l W_{il}^{(k)}\widetilde{H}_{lj}^{(k)}$ into the above inequality, we have

$$-\log\left(\sum_l W_{il}^{(k)} H_{lj}^{(k)}\right) \leq -\sum_l \frac{W_{il}^{(k)}\widetilde{H}_{lj}^{(k)}}{\sum_l W_{il}^{(k)}\widetilde{H}_{lj}^{(k)}}\left(\log W_{il}^{(k)} H_{lj}^{(k)} - \log\frac{W_{il}^{(k)}\widetilde{H}_{lj}^{(k)}}{\sum_l W_{il}^{(k)}\widetilde{H}_{lj}^{(k)}}\right). \tag{3.14}$$

Therefore, $G^{(k)}(H^{(k)}, \widetilde{H}^{(k)}) \geq F^{(k)}(H^{(k)})$. This concludes the theorem immediately. □

Obviously, the function $G(H, \widetilde{H}) = \sum_k G^{(k)}(H^{(k)}, \widetilde{H}^{(k)})$ is also an auxiliary function for the entire constraint NMF criterion $F(H) = \sum_k F^{(k)}(H^{(k)})$. Lemma 3.2 indicates that $F(H)$ is nonincreasing under the update rule (3.11). Let $\partial G(H, \widetilde{H})/\partial H_{ij}^{(k)} = 0$ and we have

$$\begin{aligned}\frac{\partial G(H, \widetilde{H})}{\partial H_{ij}^{(k)}} &= \frac{\partial G^{(k)}(H^{(k)}, \widetilde{H}^{(k)})}{\partial H_{ij}^{(k)}} \\ &= -\sum_l V_{lj}^{(k)} \frac{W_{li}^{(k)}\widetilde{H}_{ij}^{(k)}}{\sum_l W_{li}^{(k)}\widetilde{H}_{ij}^{(k)}} \frac{1}{H_{ij}^{(k)}} + \sum_l W_{li}^{(k)} + 2\alpha\left(H_{ij}^{(k)} - U_i^{(k)}\right) - \frac{\beta}{n_k}\left(\frac{1}{n_k} H_{ij}^{(k)} + U_i^{(k)}\right) \\ &= 0.\end{aligned} \tag{3.15}$$

From the above equation, it directly induces the iterative formula (3.7), and lemma 3.2 demonstrates that (3.7) converges to a local minimum. For update rule (3.5)-(3.6), the proof is similar to that of update rule (3.7) using the following auxiliary function with fixed H:

$$\begin{aligned}G(W, \widetilde{W}) &= \sum_k G^{(k)}(W^{(k)}, \widetilde{W}^{(k)}) \\ &= \sum_{ijk}\left(V_{ij}^{(k)} \log V_{ij}^{(k)} - V_{ij}^{(k)} + \left(W^{(k)} H^{(k)}\right)_{ij}\right) \\ &\quad -\sum_{ijkl} V_{ij}^{(k)} \frac{\widetilde{W}_{il}^{(k)} H_{lj}^{(k)}}{\sum_l \widetilde{W}_{il}^{(k)} H_{lj}^{(k)}}\left(\log\left(W_{il}^{(k)} H_{lj}^{(k)}\right) - \log\frac{\widetilde{W}_{il}^{(k)} H_{lj}^{(k)}}{\sum_l \widetilde{W}_{il}^{(k)} H_{lj}^{(k)}}\right) \\ &\quad +\sum_k\left(\alpha\,\mathrm{tr}\left(S_w^{(k)}\right) - \beta\left\|U^{(k)}\right\|_2^2\right).\end{aligned} \tag{3.16}$$

3.3. Incremental learning

From the above analysis, our incremental learning algorithm is designed as follows:

(i) *Sample incremental learning*. As a new training sample x_0 of the ith class is added to training set, we denote that $\widetilde{V}^{(i)} = [V^{(i)}, x_0]$. Thus the training image matrix becomes

$$\widetilde{V} = \left[V^{(1)} \ \cdots \ \widetilde{V}^{(i)} \ \cdots \ V^{(c)}\right]. \tag{3.17}$$

In this case, it only needs to perform CNMF on matrix $\widetilde{V}^{(i)}$, that is, $\widetilde{V}^{(i)} \overset{\text{CNMF}}{\approx} \widetilde{W}^{(i)}\widetilde{H}^{(i)}$. The rest decompositions such as $V^{(k)} \overset{\text{CNMF}}{\approx} W^{(k)}H^{(k)}(k \neq i)$ need not implement repetitive computation. So, sample incremental learning can be performed as follows:

$$\widetilde{V} \overset{\text{INMF}}{\approx} \widetilde{W}\widetilde{H} = \left[W^{(1)} \ \cdots \ \widetilde{W}^{(i)} \ \cdots \ W^{(c)}\right] \begin{bmatrix} H^{(1)} & & & & \\ & \ddots & & & \\ & & \widetilde{H}^{(i)} & & \\ & & & \ddots & \\ & & & & H^{(c)} \end{bmatrix}. \tag{3.18}$$

(ii) *Class incremental learning*. As a new class, denoted by matrix $V^{(c+1)}$, is added to the current training set, it forms a new training image matrix as

$$\widetilde{V} = \left[V^{(1)} \ \cdots \ V^{(c)} \ \mid \ V^{(c+1)}\right]. \tag{3.19}$$

The incremental learning settings are similar to the first item (i) that all decompositions $V^{(k)} \overset{\text{CNMF}}{\approx} W^{(k)}H^{(k)}$ $(k = 1, 2, \dots, c)$ need not compute again. We only need perform CNMF on the matrix $V^{(c+1)}$, that is, $V^{(c+1)} \overset{\text{CNMF}}{\approx} W^{(c+1)}H^{(c+1)}$. Hence, class incremental learning can be implemented as below:

$$\widetilde{V} \overset{\text{INMF}}{\approx} \widetilde{W}\widetilde{H} = \left[W^{(1)} \ \cdots \ W^{(c)} \ \mid \ W^{(c+1)}\right] \begin{bmatrix} H^{(1)} & & & \\ & \ddots & & \\ & & H^{(c)} & \\ & & & H^{(c+1)} \end{bmatrix}. \tag{3.20}$$

3.4. INMF algorithm design

Based on the above discussions, this subsection will give a detail design on our INMF algorithm for face recognition. The algorithm involves two stages, namely training stage and testing stage. Details are as follows.

Training stage

Step 1. Perform CNMF (3.9) on matrices $(V^{(i)})_{m\times n_0}$, $i = 1, 2, \dots, c$, namely,

$$\left(V^{(i)}\right)_{m\times n_0} \overset{\text{CNMF}}{\approx} \left(W^{(i)}\right)_{m\times r_0}\left(H^{(i)}\right)_{r_0\times n_0}, \quad i = 1, 2, \dots, c. \tag{3.21}$$

Step 2. INMF is obtained as

$$V_{m\times n} \overset{\text{INMF}}{\approx} W_{m\times r}H_{r\times n}, \tag{3.22}$$

where $r = cr_0$, $n = cn_0$, and

$$W_{m\times r} = \left[W^{(1)} \ W^{(2)} \ \cdots \ W^{(c)}\right], \qquad H_{r\times n} = \operatorname{diag}\left(H^{(1)}, H^{(2)}, \dots, H^{(c)}\right). \tag{3.23}$$

If there is a new training sample or class added to current training set, then the incremental learning algorithm presented in Section 3.4 is applied to this stage.

Recognition stage

Step 3. Calculate the coordinates of a testing sample $\widehat{v}$ in the feature space $\operatorname{span}\{W_1, W_2, \dots, W_r\}$ by $\widehat{h} = W^{+}\widehat{v}$, where W^{+} is the Moore-Penrose inverse of W.

Step 4. Compute the mean column vector $\overline{v}_i$ of class i and its coordinates vector $h_i = W^{+}\overline{v}_i$ $(i = 1, 2, \dots, c)$. The testing image $\widehat{v}$ is classified to class k, if $d(\widehat{h}, h_k) = \min_{1 \le i \le c} d(\widehat{h}, h_i)$, where $d(\widehat{h}, h_i)$ denotes the Euclidean distance between vectors $\widehat{h}$ and h_i.

3.5. Sparseness of coefficient matrix H

Let $h \in \mathbb{R}^n$, define sparseness function with L^1 and L^2 norms [7] by

$$f_{\text{sparse}}(h) = \frac{\sqrt{n} - \|h\|_1 / \|h\|_2}{\sqrt{n} - 1}. \tag{3.24}$$

It can be seen that sparseness function $f_{\text{sparse}} : \mathbb{R}^n \to \mathbb{R}$ with range $[0, 1]$.

For INMF method, we have the following theorem for each column h_i of H.

Theorem 3.4. *Sparseness of each column h_i of H in INMF has the following estimation:*

$$\frac{\sqrt{cr_0} - \sqrt{r_0}}{\sqrt{cr_0} - 1} \le f_{\{\text{sparse}\}}(h_i) \le 1. \tag{3.25}$$

Proof. Let

$$h_i = \left(0, \dots, 0, h_{i1}^{(j)}, \dots, h_{ir_0}^{(j)}, 0, \dots, 0\right)^T \in \mathbb{R}^r, \qquad \widetilde{h}_i = \left(h_{i1}^{(j)}, \cdots, h_{ir_0}^{(j)}\right)^T \in \mathbb{R}^{r_0}, \tag{3.26}$$

where h_i belongs to class i in H.

Obviously,

$$\|h_i\|_1 = \|\widetilde{h}_i\|_1, \qquad \|h_i\|_2 = \|\widetilde{h}_i\|_2. \tag{3.27}$$

Moreover,

$$1 \le \frac{\|\widetilde{h}\|_1}{\|\widetilde{h}\|_2} \le \sqrt{r_0}. \tag{3.28}$$

So, we have

$$\frac{\sqrt{r} - \sqrt{r_0}}{\sqrt{r} - 1} \le \frac{\sqrt{r} - \|h_i\|_1 / \|h_i\|_2}{\sqrt{r} - 1} \le 1. \tag{3.29}$$

It concludes for $r = cr_0$ that

$$\frac{\sqrt{cr_0} - \sqrt{r_0}}{\sqrt{cr_0} - 1} \le f_{\text{sparse}}(h_i) \le 1. \tag{3.30}$$

□

In the experimental section, the parameters are selected as $r_0 = 4$ and $c = 120$ using INMF on FERET database. It can be calculated that

$$0.9522 \le f_{\text{sparse}}(h_i) \le 1. \tag{3.31}$$

While on CMU PIE database, we select $r_0 = 4$ and $c = 68$ and calculate that

$$0.9355 \le f_{\text{sparse}}(h_i) \le 1. \tag{3.32}$$

These demonstrate that each column of H in INMF is highly sparse. Apparently, the coefficient column vectors between different classes in H are automatically orthogonal.

3.6. Computational complexity

This section discusses the computational complexity of our proposed INMF approach. The ith iterative procedure of proposed INMF includes two parts, namely $W^{(i)}$ and $H^{(i)}$. For each matrix $V^{(i)}$ the iteration for $W^{(i)}$ needs $mr_0(n_0r_0 + 2n_0 + 2)$ multiple times. While for $H^{(i)}$, it needs $n_0r_0(mr_0 + 2m + 10)$ multiple times. Therefore, the total running multiple times of our INMF are

$$T_{\text{INMF}} = \left(2mn_0r_0^2 + 4mn_0r_0 + 2mr_0 + 10n_0r_0\right)c = \frac{2mnr^2}{c^2} + \frac{4mnr}{c} + \frac{10nr}{c} + 2mr. \qquad (3.33)$$

Similar to INMF, we can obtain the running multiple times of NMF approach as $T_{\text{NMF}} = 2mnr^2 + 4mnr + 2mr + 2nr$. It can be seen that the computational complexity of our INMF method is greatly lower than that of NMF.

4. Experimental results

In this section, FERET and CMU PIE databases are selected to evaluate the performance of our INMF method along with BNMF, NMF, and PCA methods. All images in two databases are aligned by the centers of eyes and mouth and then normalized with resolution 112 × 92. The original images with resolution 112 × 92 are reduced to wavelet feature face with resolution 30 × 25 after two-level D4 wavelet decomposition. If there are negative pixels in the wavelet faces, we will transform them into nonnegative faces with simple translations. The nearest neighbor classifier using Euclidean distance is exploited here. In the following experiments, the parameters are set to $r = 120$ for NMF, $r_0 = 4$ for BNMF and INMF, $\alpha = 10^{-4}$, $\beta = 10^{-3}$ for INMF. The stopping condition of iterative update is

$$\frac{F^{(n-1)} - F^{(n)}}{F^{(n)}} \leq \delta, \qquad (4.1)$$

where $F^{(n)}$ is the nth update criterion function defined in (3.3), the threshold δ is set to 10^{-12}. We stop the iteration if stopping condition (4.1) is met or if exceeding 1000 times iteration.

4.1. Face databases

In FERET database, we select 120 people, 6 images for each individual. The six images are extracted from 4 different sets, namely Fa, Fb, Fc, and duplicate. Fa and Fb are sets of images taken with the same camera at the same day but with different facial expressions. Fc is a set of images taken with different camera at the same day. Duplicate is a set of images taken around 6–12 months after the day taking the Fa and Fb photos. Details of the characteristics of each set can be found in [3]. Images from one individual are shown in Figure 1.

CMU PIE database includes totally 68 people. There are 13 pose variations ranging from full right-profile image to full left-profile image and 43 different lighting conditions, 21 flashes with ambient light on or off. In our experiment, for each person, we select 56 images including 13 poses with neutral expression and 43 different lighting conditions in frontal view. Part images of one person are shown in Figure 2.

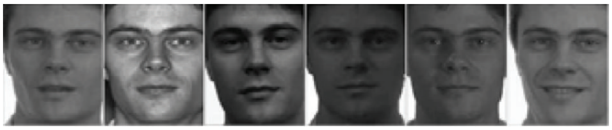


Figure 1: Images of one person from FERET database.

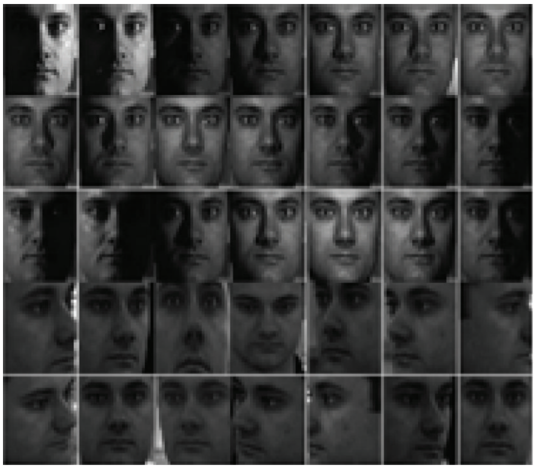


Figure 2: Part images of one person from CMU PIE database.

4.2. Basis face images

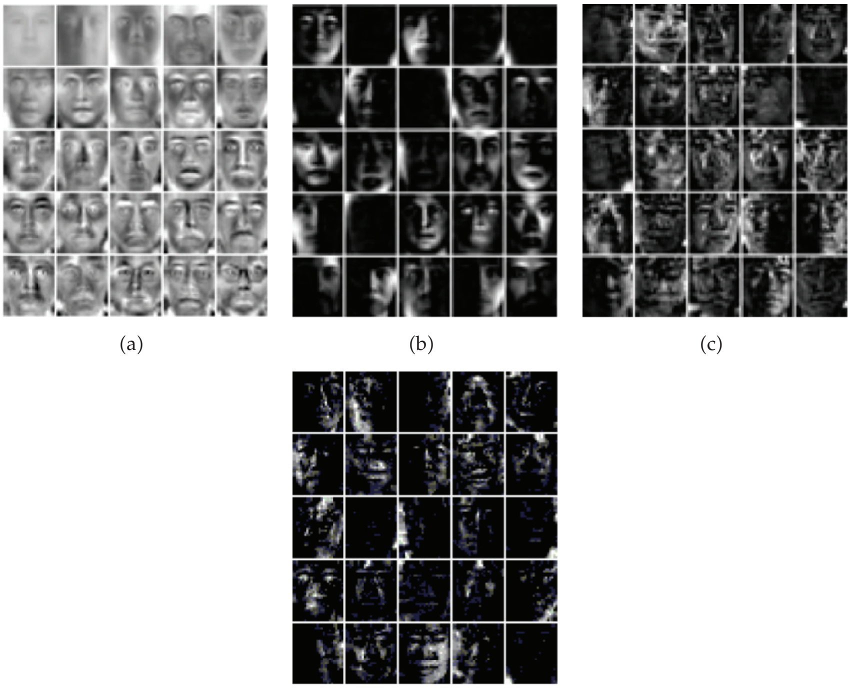
This section shows the basis images of the training set learnt by PCA, NMF, BNMF, and INMF approaches. Figure 3 shows 25 basis images of each approach on CMU PIE database. It can be seen that the bases of all methods are additive except for PCA. PCA extracts the holistic facial features. INMF learns more local features than NMF and BNMF. Moreover, the greater number of basis image is, the more localization is learnt in all NMF-based approaches.

4.3. Results on FERET database

This section reports the experimental results with nonincremental learning and incremental learning on FERET database. All methods use the same training and testing face images. The experiments are repeated 10 times; and the average accuracies under different training number, along with the mean running times, are recorded.

4.3.1. Nonincremental learning

We randomly select n ($n = 2, 3, 4, 5$) images from each person for training, while the rest of ($6-n$) images of each individual for testing. The average accuracies of training samples ranging from 2 to 5 are recorded in Table 1 and plotted in Figure 4(a). The recognition accuracies of INMF, BNMF, NMF, and PCA are 66.73%, 66.07%, 64.44%, and 34.33%, respectively, with 2 training images. The performance for each method is improved when the number of training images increases. When the number of training images is equal to 5, the recognition accuracies of INMF, BNMF, NMF, and PCA are 83.08%, 81.67%, 80.25%, and 37.58%, respectively.



(a) (b) (c)

(d)

Figure 3: Comparisons on basis images of PCA, NMF, BNMF, and INMF (from left to right), respectively, on CMU PIE database results.

In addition, Table 2 gives the comparisons on average time-consuming in three NMF-based approaches. It can be seen that our INMF method gives the best performance for all cases of nonincremental learning on FERET database.

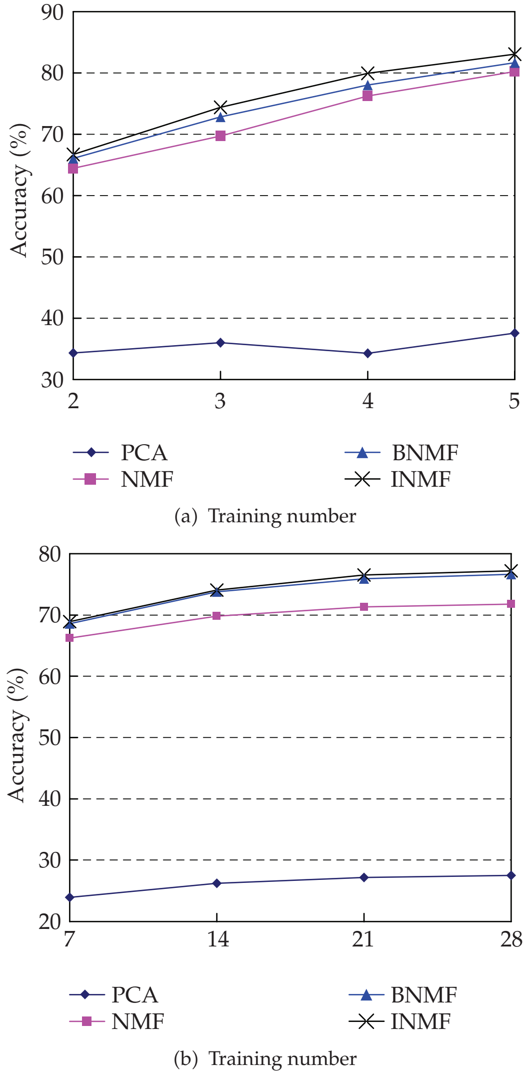
4.3.2. Class incremental learning

For 119 people, we randomly select 3 images from each individual for training and then add a new class to the training set. NMF must conduct repeated learning while BNMF and INMF need merely perform incremental training on the new added class. The average accuracies and the mean running times are recorded in Table 3 (plotted in Figure 6(a)) and Table 4, respectively.

Compared with the NMF and BNMF approaches, the proposed method gives around 5% and 1.5% accuracy improvements, respectively. The running time of INMF is around 2 times and 219 times faster than that of NMF with 119 and 120 individuals for training and class-incremental learning, respectively. Above all, our INMF gives the best performance on FERET database.

4.4. Results on CMU PIE database

The experimental setting on CMU PIE database is similar to that of FERET database. It also includes two parts, namely nonincremental training and incremental learning. The



(a) Training number

(b) Training number

Figure 4: Accuracy comparisons on (a) FERET and (b) CMU PIE databases.

Table 1: Accuracy comparisons on FERET database.

Training number	PCA	NMF	BNMF	INMF
2	34.33%	64.44%	66.07%	66.73%
3	36.00%	69.72%	72.81%	74.39%
4	34.29%	76.25%	78.04%	78.92%
5	37.58%	80.25%	81.67%	83.08%

experiments are repeated 10 times and the average accuracies under different training number, along with the mean running times, are recorded for comparisons. Details are as follows.

Table 2: Running times (s) on FERET database.

Training number	NMF(s)	BNMF(s)	INMF(s)
2	51.14	30.69	21.12
3	74.58	57.02	33.68
4	100.12	75.81	45.34
5	122.61	89.34	56.30

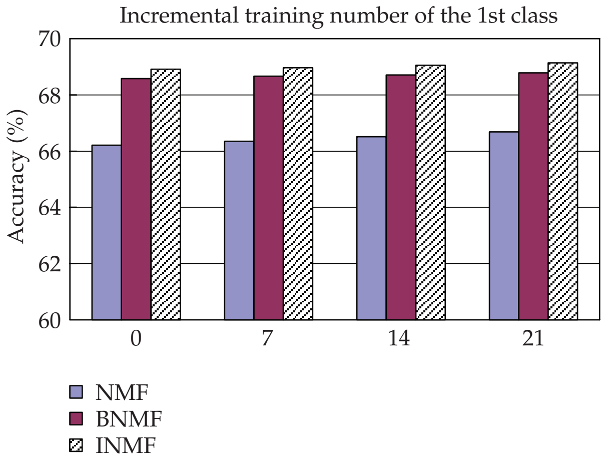


Figure 5: Comparisons on sample incremental learning.

Table 3: Accuracy comparisons on class incremental learning.

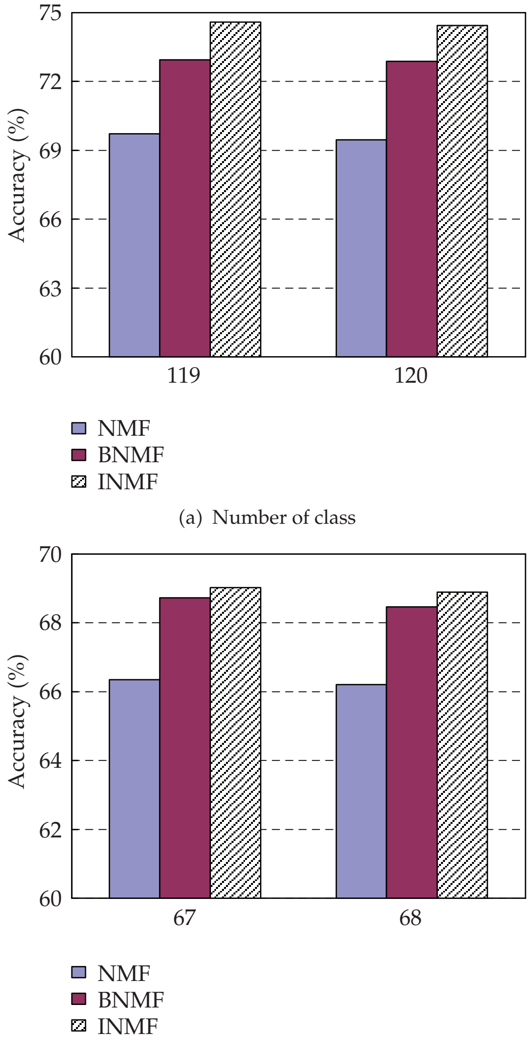
Number of class	NMF	BNMF	INMF
119	69.72%	72.94%	74.59%
120	69.45%	72.88%	74.44%

4.4.1. Nonincremental learning

For each individual, $n\,(n = 7, 14, 21, 28)$ images are randomly selected for training, while the rest $(56 - n)$ images for testing. The average recognition rates and mean running times are tabulated in Table 5 (plotted in Figure 4(b)) and Table 6, respectively. It can be seen that the recognition accuracies of INMF, BNMF, NMF, and PCA are 68.91%, 68.58%, 66.21%, and 23.94%, respectively with training number 7. When the number of training images is equal to 28, the recognition accuracies of INMF, BNMF, NMF, and PCA are 77.18%, 76.64%, 71.77%, and 27.51%, respectively. Compared with the PCA and NMF methods, the proposed method gives around 49% and 5% accuracy improvements, respectively. The performance of INMF is slightly better than that of BNMF. However, the computational efficiency of INMF greatly outperforms BNMF.

4.4.2. Sample incremental learning

We randomly select 7 images from each person for training, and the rest 49 images for testing. In the sample-incremental learning stage, 7, 14, and 21 images of the first individual are added



(a) Number of class

(b) Number of class

Figure 6: Class incremental learning comparisons on (a) FERET and (b) CUM PIE databases.

to the training set, respectively, while the training images from the rest individuals are kept unchanged. Table 7 (Figure 5) and Table 8 show the average recognition accuracies and the mean running times, respectively. Experimental results show that our INMF method gives the best performance for all cases.

4.4.3. Class incremental learning

For 67 people, we randomly select 7 images from each individual for training and then add a new class to the training set. NMF should conduct repetitive learning. BNMF and INMF need merely to perform incremental learning on the new added class. The average recognition rates and the mean running times are recorded in Table 9 (plotted in Figure 6(b)) and Table 10,

Table 4: Running times (s) on class incremental learning.

Number of class	NMF(s)	BNMF(s)	INMF(s)
119	72.57	56.12	32.34
120	74.58	0.86	0.34

Table 5: Accuracy comparisons on CMU PIE database.

Training number	PCA	NMF	BNMF	INMF
7	23.94%	66.21%	68.58%	68.91%
14	26.24%	69.82%	73.79%	74.07%
21	27.15%	71.33%	75.93%	76.55%
28	27.51%	71.77%	76.64%	77.18%

Table 6: Running times (s) on CMU PIE database.

Training number	NMF(s)	BNMF(s)	INMF(s)
7	79.58	57.35	37.24
14	144.52	114.61	89.50
21	208.20	164.18	124.45
28	267.97	215.94	176.23

Table 7: Accuracy comparisons on sample incremental learning.

Incremental training number	NMF	BNMF	INMF
0	66.21%	68.58%	68.91%
7	66.35%	68.66%	68.97%
14	66.51%	68.71%	69.05%
21	66.68%	68.78%	69.14%

Table 8: Running times on sample incremental learning.

Incremental training number	NMF (s)	BNMF (s)	INMF (s)
0	79.58	57.35	37.24
7	79.77	2.34	1.20
14	80.02	3.12	2.36
21	80.23	4.53	3.21

Table 9: Comparisons on class incremental learning.

Number of class	NMF	BNMF	INMF
67	66.35%	68.73%	69.02%
68	66.21%	68.46%	68.89%

respectively. Experimental results show that INMF outperforms BNMF and NMF in both recognition rates and computational efficiency.

Table 10: Running time (s) on class incremental learning.

Number of class	NMF(s)	BNMF(s)	INMF(s)
67	72.45	55.19	34.28
68	79.58	1.36	0.70

5. Conclusions

This paper proposed a novel constraint INMF method to address the time-consuming problem and incremental learning problem of existing NMF-based approaches for face recognition. INMF has some good properties, such as low computational complexity; sparse coefficient matrix; orthogonal coefficient column vectors between different classes in coefficient matrix H; especially for incremental learning, and so on. Experimental results on FERET and CMU PIE face database show that INMF outperforms PCA, NMF, and BNMF approaches in nonincremental learning and incremental learning.

Acknowledgments

This work is supported by NSF of China (60573125, 60603028) and in part by the Program for New Century Excellent Talents of Educational Ministry of China (NCET-06-0762), NSF of Chongqing CSTC (CSTC2007BA2003). The authors would like to thank the US Army Research Laboratory for contribution of the FERET database and CMU for the CMU PIE database.

References

[1] R. Chellappa, C. L. Wilson, and S. Sirohey, “Human and machine recognition of faces: a survey,” *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–741, 1995.

[2] W. Zhao, R. Chellappa, A. Rosenfeld, and J. Phillips, “Face recognition: a literature survey,” Tech. Rep. CFAR-TR00-948, University of Maryland, College Park, Md, USA, 2000.

[3] R. Brunelli and T. Poggio, “Face recognition: features versus templates,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1052, 1993.

[4] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[5] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Proceedings of the Advances in Neural Information Processing Systems (NIPS ’01)*, vol. 13, pp. 556–562, Vancouver, Canada, December 2001.

[6] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[7] S. Z. Li, X. W. Hou, H. J. Zhang, and Q. S. Cheng, “Learning spatially localized, parts-based representation,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR ’01)*, vol. 1, pp. 207–212, Kauai, Hawaii, USA, December 2001.

[8] S. Wild, J. Curry, and A. Dougherty, “Improving non-negative matrix factorizations through structured initialization,” *Pattern Recognition*, vol. 37, no. 11, pp. 2217–2232, 2004.

[9] I. Buciu and I. Pitas, “A new sparse image representation algorithm applied to facial expression recognition,” in *Proceedings of the 14th IEEE Workshop on Machine Learning for Signal Processing (MLSP ’04)*, pp. 539–548, Sao Luis, Brazil, September-October 2004.

[10] P. O. Hoyer, “Non-negative matrix factorization with sparseness constraints,” *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[11] Y. Xue, C. S. Tong, W.-S. Chen, and W. Zhang, “A modified non-negative matrix factorization algorithm for face recognition,” in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR ’06)*, vol. 3, pp. 495–498, Hong Kong, August 2006.

[12] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas, "Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 683–695, 2006.

[13] I. Buciu and I. Pitas, "NMF, LNMF, and DNMF modeling of neural receptive fields involved in human facial expression perception," *Journal of Visual Communication and Image Representation*, vol. 17, no. 5, pp. 958–969, 2006.

[14] A. Pascual-Montano, J. M. Carazo, K. Kochi, D. Lehmann, and R. D. Pascual-Marqui, "Nonsmooth nonnegative matrix factorization (nsNMF)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 403–415, 2006.

[15] I. Kotsia, S. Zafeiriou, and I. Pitas, "A novel discriminant non-negative matrix factorization algorithm with applications to facial image characterization problems," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 588–595, 2007.

[16] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Computation*, vol. 19, no. 10, pp. 2756–2779, 2007.

[17] C.-J. Lin, "On the convergence of multiplicative update algorithms for nonnegative matrix factorization," *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1589–1596, 2007.

[18] T. Zhang, B. Fang, Y. Y. Tang, G. He, and J. Wen, "Topology preserving non-negative matrix factorization for face recognition," *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 574–584, 2008.

[19] B. B. Pan, W. S. Chen, and C. Xu, "Incremental learning of face recognition based on block non-negative matrix factorization," 2008 (Chinese), to appear in *Computer Application Research*.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 825215, 15 pages
doi:10.1155/2008/825215

Research Article

Direct Neighborhood Discriminant Analysis for Face Recognition

Miao Cheng, Bin Fang, Yuan Yan Tang, and Jing Wen

Department of Computer Science, Chongqing University, Chongqing 400030, China

Correspondence should be addressed to Bin Fang, fb@cqu.edu.cn

Received 21 March 2008; Accepted 16 April 2008

Recommended by Ming Li

Face recognition is a challenging problem in computer vision and pattern recognition. Recently, many local geometrical structure-based techiniques are presented to obtain the low-dimensional representation of face images with enhanced discriminatory power. However, these methods suffer from the small simple size (SSS) problem or the high computation complexity of high-dimensional data. To overcome these problems, we propose a novel local manifold structure learning method for face recognition, named direct neighborhood discriminant analysis (DNDA), which separates the nearby samples of interclass and preserves the local within-class geometry in two steps, respectively. In addition, the PCA preprocessing to reduce dimension to a large extent is not needed in DNDA avoiding loss of discriminative information. Experiments conducted on ORL, Yale, and UMIST face databases show the effectiveness of the proposed method.

1. Introduction

Many pattern recognition and data mining problems involve data in very high-dimensional spaces. In the past few decades, face recognition (FR) has become one of the most active topics in machine vision and pattern recognition, where the feature dimension of data usually can be very large and hardly handled directly. To get a high recognition rate for FR, numerous feature extraction and dimension reduction methods have been proposed to find the low-dimensional feature representation with enhanced discriminatory power. Among these methods, two state-of-the-art FR methods, principle component analysis (PCA) [1], and linear discriminant analysis (LDA) [2] have been proved to be useful tools for dimensionality reduction and feature extraction.

LDA is a popular supervised feature extraction technique for pattern recognition, which intends to find a set of projective direction to maximize the between-class scatter matrix S_b and minimize the within-class scatter matrix S_w simultaneously. Although successful in many cases, many LDA-based algorithms suffer from the so-called "small sample size" (SSS) problem that exists when the number of available samples is much smaller than the dimensionality of the samples, which is particularly problematic in FR applications. To solve this problem, many extensions of LDA have been developed in the past. Generally, these approaches to address SSS problem can be divided into three categories, namely, Fisherface method, Regularization methods, and Subspace methods. Fisherface methods incorporate a PCA step into the LDA framework as a preprocessing step. Then LDA is performed in the lower dimensional PCA subspace [2], where the within-class scatter matrix is no longer singular. Regularization methods [3, 4] add a scaled identity matrix to scatter matrix so that the perturbed scatter matrix becomes nonsingular. However, Chen et al. [5] have proved that the null space of S_w contains the most discriminate information, while the SSS problem takes place, and proposed the null space LDA (NLDA) method which only extracts the discriminant features present in the null space of the S_w. Later, Yu and Yang [6] utilized discriminatory information of both S_b and S_w, and proposed a direct-LDA (DLDA) method to solve SSS problem.

Recently, the motivation for finding the manifold structure in high-dimensionality data elevates the wide application of manifold learning in data mining and machine learning. Among these methods, Isomap [7], LLE [8], and Laplacian eigenmaps [9, 10] are representative techniques. Based on the locality preserving concept, some excellent local embedding analysis techniques are proposed to find the manifold structure based on local nearby data [11, 12]. However, these methods are designed to preserve the local geometrical structure of original high-dimensional data in the lower dimensional space rather than good discrimination ability. In order to get a better classification effect, some supervised learning techniques are proposed by incorporating the discriminant information into the locality preserve learning techniques [13–15]. Moreover, Yan et al. [15] explain the manifold learning techniques and the traditional dimensionality reduction methods as a unified framework that can be defined in a graph embedding way instead of a kernel view [16]. However, the SSS problem is still exists in the graph embedding-based discriminant techniques. To deal with such problem, PCA is usually performed to reduce dimension as a preprocessing step in such environment [11, 15].

In this paper, we present a two-stage feature extraction technique named direct neighborhood discriminant analysis (DNDA). Compared to other geometrical structure learning work, the PCA step is not needed to be done in our method. Thus, more discriminant information can be kept for FR purpose, and as a result improved performance is expected. The rest of the paper is structured as follows: we give a brief review of LDA and DLDA in Section 2. We then introduce in Section 3 the proposed method for dimensionality reduction and feature extraction in FR. The effectiveness of our method is evaluated in a set of FR experiments in Section 4. Finally, we give concluding remarks in Section 5.

2. Review of LDA and DLDA

2.1. LDA

LDA is a very popular technique for linear feature extraction and dimensionality reduction [2], which chooses the basis vectors of the transformed space as those directions of the original space to make the ratio of the between-class scatter and the within-class scatter are maximized. Formally, the goal of LDA is to seek the optimal orthogonal matrix w, such that maximizing the following quotient, the Fisher Criterion:

$$J(W) = \arg\max_{w} \frac{w^T S_b w}{w^T S_w w}, \tag{2.1}$$

where S_b is the between-class scatter matrix, S_w is the within-class scatter matrix, such that w can be formed by the set of generalized eigenvectors corresponding to following eigenanalysis problem:

$$S_b w = \lambda S_w w. \tag{2.2}$$

When the inverse of S_w exists, the generalize vectors can be obtained by eigenvalue decomposition of $S_w^{-1} S_b$. However, one usually confronts the difficulty that the within-class scatter matrix S_w is singular (SSS) in FR problem. The so-called PCA plus LDA approach [2] is a very popular technique which intends to overcome such circumstances.

2.2. DLDA

To take discriminant information of both S_b and S_w into account without conducting PCA, a direct LDA (DLDA) technique has been presented by Yu and Yang [6]. The basic idea behind the approach is that no significant information will be lost if the null space of S_b is discarded. Based on the assumption, it can be concluded that the optimal discriminant features exist in the range space of S_b.

Let multiclass classification be considered, given a data matrix $X \in R^{d \times N}$, where each column x_i represents a sample data. Suppose X is composed of c classes and total number of samples is denoted by $\sum_{i=1}^{c} N_i = N$, for the ith class consists of N_i samples. Then, the between-class scatter matrix is defined as

$$S_b = \frac{1}{N} \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T = G_b G_b^T, \tag{2.3}$$

where

$$G_b = \left[\frac{\sqrt{N_1}}{\sqrt{N}} (\mu_1 - \mu), \frac{\sqrt{N_2}}{\sqrt{N}} (\mu_2 - \mu), \ldots, \frac{\sqrt{N_c}}{\sqrt{N}} (\mu_c - \mu) \right], \tag{2.4}$$

$$\mu_i = \frac{1}{N_i} \sum_{m=1}^{N_i} x_m \tag{2.5}$$

are the class mean sample, and

$$\mu = \frac{1}{N} \sum_{i=1}^{c} N_i \mu_i \tag{2.6}$$

denotes the total mean sample. Similarly, the within-class scatter matrix is defined as

$$S_w = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_j - \mu_i)(x_j - \mu_i)^T = G_w G_w^T, \tag{2.7}$$

where,

$$G_w = \left[\frac{1}{\sqrt{N}}(x_1 - \mu_{c_1}), \frac{1}{\sqrt{N}}(x_2 - \mu_{c_2}), \ldots, \frac{1}{\sqrt{N}}(x_N - \mu_{c_N})\right]. \tag{2.8}$$

In DLDA, eigenvalue decomposition is performed on the between-class matrix S_b, firstly. Suppose the rank of S_b is t, and let $D_b = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_t)$ be a diagonal matrix with the t largest eigenvalue on the main diagonal in descending order, $Y = [v_1, v_2, \ldots, v_t]$ is the eigenvector matrix that consists of t corresponding eigenvectors. Then, dimensionality of data x is reduced by using the projection matrix $Z = YD_b^{-1/2}$ from d to t, $Z^T x$. And eigenvalue decomposition is performed on the within-class scatter matrix of the projected samples, $\widetilde{S_w} = Z^T S_w Z$. Let $D_w = \mathrm{diag}(\eta_1, \eta_2, \ldots, \eta_t)$ be the ascending order eigenvalue matrix of $\widetilde{S_w}$ and $U = [u_1, u_2, \ldots, u_t]$ be the corresponding eigenvector matrix. Therefore, the final transformation matrix is given by $W = YD_b^{-1/2} U D_w^{-1/2}$.

To address the computation complexity problem of high dimensional data, the eigenanalysis method presented by Turk and Pentland [1] is applied in DLDA, which makes the eigenanalysis of scatter matrices be progressed in an efficient way. For the eigenvalue decomposition of any symmetry matrix A with the form of $A = BB^T$, we can consider the eigenvectors v_i of $B^T B$ such that

$$B^T B v_i = \lambda_i v_i. \tag{2.9}$$

Premultiplying both sides by G, we have

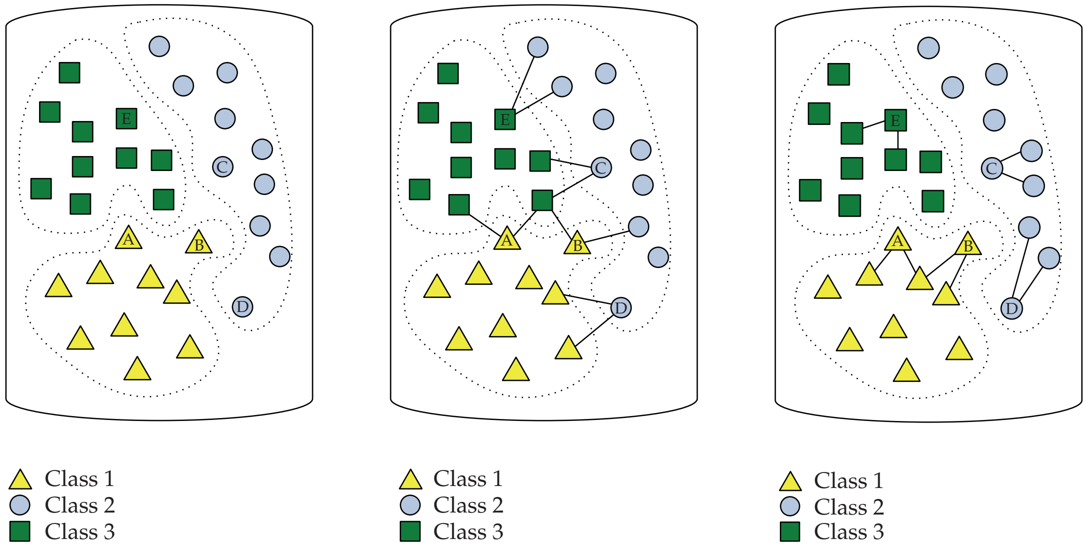
$$BB^T B v_i = AB v_i = \lambda_i B v_i \tag{2.10}$$

from which it can be concluded that the eigenvectors of A is Bv_i with the corresponding eigenvalue λ_i.

3. Direct neighborhood discriminant analysis

Instead of mining the statistical discriminant information, manifold learning techniques try to find out the local manifold structure of data. Derived from the locality preserving idea [10, 11], graph embedding framework-based techniques extract the local discriminant features for classification. For a general pattern classification problem, it is expected to find a linear transformation, such that the compactness for the samples that belong to the same class and the separation for the samples of the interclass should be enhanced in the transformed space. As an example, a simple multiclass classification problem is illustrated in Figure 1. Suppose there are two nearest inter- and intraclass neighbors searched for classification. The inter- and intracalss nearby data points of five data points A–E is shown in Figures 1(b) and 1(c), respectively. For data point A, it is optimal that the distance from its interclass neighbors should be maximized to alleviate their bad influence for classification. On the other hand, the distance between data point A and its intraclass neighbors should be minimized to make A be classified correctly.

Based on the consideration, two graphs, that is, the between-class graph G and the within-class graph G' are constructed to discover the local discriminant structure [13, 15]. For each data point x_i, its sets of inter- and intraclass neighbors are indicated by $kNN^b(x_i)$ and



(a) (b) (c)

Figure 1: Local discriminant neighbors. (a) Multi-class classification (b) Two interclass neighbors (c) Two intraclass neighbors.

$kNN^w(x_i)$, respectively. Then, the weight W_{ij} reflects the weight of the edge in the between-class graph G is defined as

$$W_{ij}^b = \begin{cases} 1 & \text{if } x_i \in kNN^b(x_j) \text{ or } x_j \in kNN^b(x_i), \\ 0 & \text{else}, \end{cases} \tag{3.1}$$

and similarly define within-class affinity weight as

$$W_{ij}^w = \begin{cases} 1 & \text{if } x_i \in kNN^w(x_j) \text{ or } x_j \in kNN^w(x_i), \\ 0 & \text{else}. \end{cases} \tag{3.2}$$

Let the transformation matrix be denoted by $P \in R^{d \times d'}(d' \ll d)$, which transforms the original data x from high-dimensional space R^d into a low-dimensional space $R^{d'}$ by $y = P^T x$. The separability of interclass samples in the transformed low-dimensional space can be defined as

$$\begin{aligned} F_b &= \sum_{i,j} \left\| P^T x_i - P^T x_j \right\|^2 W_{ij}^b \\ &= \sum_{i,j} \operatorname{tr}\left[\left(P^T x_i - P^T x_j \right) \left(P^T x_i - P^T x_j \right)^T W_{ij}^b \right] \\ &= \sum_{i,j} \operatorname{tr}\left[P^T (x_i - x_j) W_{ij}^b (x_i - x_j)^T P \right] \\ &= \operatorname{tr}\left(2 \sum_i P^T x_i D_{ii}^b x_i^T P - 2 \sum_{i,j} P^T x_i W_{ij}^b x_j^T P \right) \\ &= \operatorname{tr}\left(P^T X \left(2D^b - 2W^b \right) X^T P \right), \end{aligned} \tag{3.3}$$

where $\mathrm{tr}(\cdot)$ is the trace of matrix, $X = [x_1, x_2, \ldots, x_N]$ is the data matrix, and D^b is a diagonal matrix, of which entries are column (or row, since W^b is symmetric) sum of W^b, $D^b_{ii} = \sum_j W^b_{ij}$. Similarly, the compactness of intraclass samples can be characterized as

$$\begin{aligned}
F_w &= \sum_{i,j} \left\| P^T x_i - P^T x_j \right\|^2 W^w_{ij} \\
&= \sum_{i,j} \mathrm{tr}\left[\left(P^T x_i - P^T x_j \right) \left(P^T x_i - P^T x_j \right)^T W^w_{ij} \right] \\
&= \sum_{i,j} \mathrm{tr}\left[P^T \left(x_i - x_j \right) W^w_{ij} \left(x_i - x_j \right)^T P \right] \\
&= \mathrm{tr}\left(2 \sum_i P^T x_i D^w_{ii} x_i^T P - 2 \sum_{i,j} P^T x_i W^w_{ij} x_j^T P \right) \\
&= \mathrm{tr}\left(P^T X \left(2D^w - 2W^w \right) X^T P \right).
\end{aligned} \tag{3.4}$$

Here, D^w is a diagonal matrix of which entries are column (or row) sum of W^w on the main diagonal, $D^w_{ii} = \sum_j W^w_{ij}$. Then, the optimal transformation matrix P can be obtained by solving the following problem:

$$\begin{aligned}
P^* &= \arg\max_P \frac{P^T SsP}{P^T S_c P}, \\
S_s &= X \left(2D^b - 2W^b \right) X^T, \\
S_c &= X \left(2D^w - 2W^w \right) X^T.
\end{aligned} \tag{3.5}$$

Here, S_c is always singular with small training sample set leading problem to get projective matrix P directly, thus previous local discriminant techniques still suffer from the curse of high dimensionality. Generally, PCA is usually performed to reduce dimension as a preprocessing step in such environment [15], however, possible discriminant information may be ignored. Inspired by DLDA, we can perform eigenanalysis on S_s and S_c successively to extract the complete local geometrical structure directly without PCA preprocessing. To alleviate the burden of computation, we reformulate S_s and S_c so that Turk's eigenanalysis method can be employed. For each nonzero element of W^b, W^b_{ij}, we build an N dimensional interclass index vector $h^{(i,j)}$ of all zeroes except the ith and jth element is set to be 1 and -1, respectively:

$$h^{(i,j)} = \left[\overbrace{0 \ \cdots \ 0}^{i-1}, \overset{i}{1}, 0 \ \cdots \ 0, \overset{j}{-1}, \overbrace{0 \ \cdots \ 0}^{N-j} \right]^T. \tag{3.6}$$

Suppose there are N_b nonzero elements in W^b, let $H_s = [h_1, h_2, \ldots, h_{N_b}]$ be the interclass index matrix made up of N_b interclass index vectors. It can be easily obtained that $2D^b - 2W^b = H_s H_s^T$, which we prove in Appendix A. Therefore, S_s can be reformulated as

$$S_s = X H_s H_s^T X^T. \tag{3.7}$$

```
Input: Data matrix $X \in R^{d \times N}$, class label $L$
Output: Transformed matrix $P^*$
1. Construct the between-class and the within-class affinity weight matrix $W^b$, $W^w$.
2. Construct the interclass and the intraclass index matrix $H_s$, $H_c$ according to the nonzero elements of $W^b$, $W^w$.
For the $k$th nonzero element of $W^b(W^w)$, $W^b_{ij}(W^w_{ij})$, the corresponding $k$th column in $H_s(H_c)$ is constructed as
    $\left[\overbrace{0 \cdots 0}^{i-1},\overset{i}{1},0 \cdots 0,\overset{j}{-1},\overbrace{0 \cdots 0}^{N-j}\right]^T.$
3. Apply eigenvalue decomposition to $S_s$ and keep the largest $t$ nonzero eigenvalues $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_t]$ and corresponding eigenvectors $U = [u_1, u_2, \ldots, u_t]$ after sorted in decreasing order, where $t = \mathrm{rank}(S_s)$.
4. Compute $P_s$ as $P_s = UD_s^{-1/2}$, where $D_s = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_t)$ is diagonal matrix with $\lambda_i$ on the main diagonal.
5. Perform eigenvalue decomposition on $\widetilde{S_c} = P_s^T S_c P_s$. Let $D_c = \mathrm{diag}(\mu_1, \mu_2, \ldots, \mu_n)$ be the eigenvalue matrix of $\widetilde{S_c}$ in ascending order and $V = [v_1, v_2, \ldots, v_n]$ be the corresponding eigenvector matrix. Calculate $P_c$ as $P_c = VD_c^{-1/2}$.
6. $P^* \leftarrow P_s P_c$.
```

Algorithm 1: DNDA algorithm.

(a)

(b)

(c)

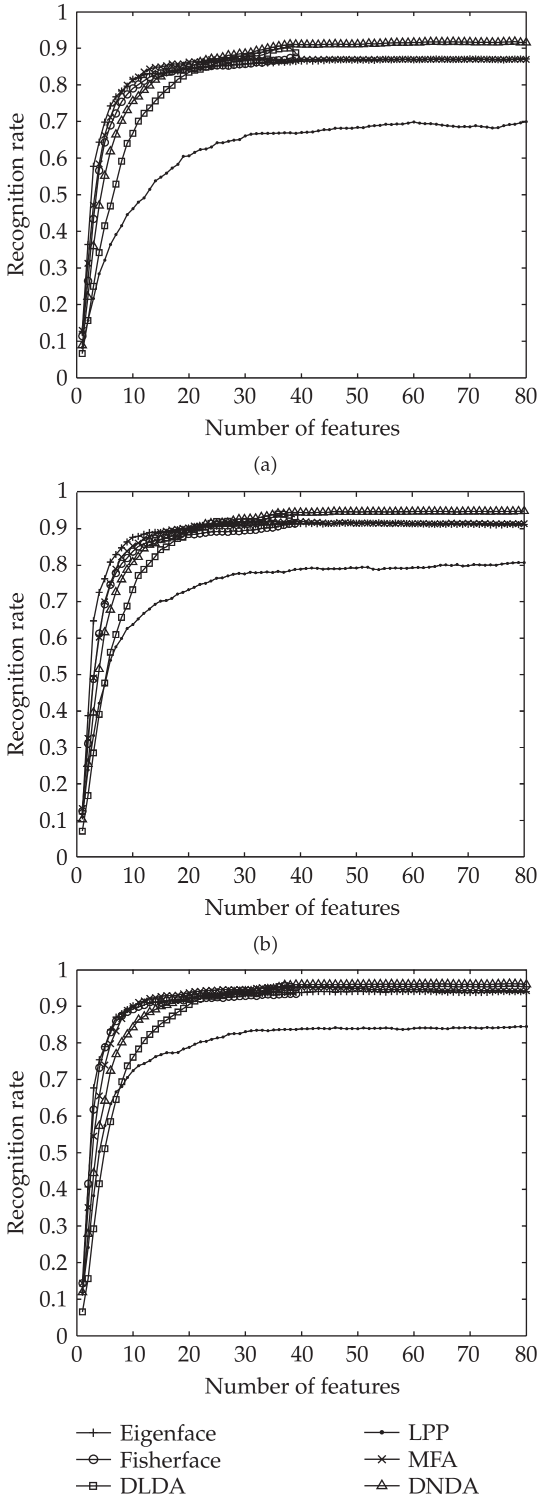
Figure 2: Sample images from ORL, Yale, and UMIST face database. (a) ORL, (b) Yale, and (c) UMIST.

As each column in H_s has only two nonzero elements 1 and −1, we can make the first row in H_s be a null row by adding all rows but the first to the first row. On the other hand, for each column $h^{(j,i)}$ in H_s, there is another column $h^{(j,i)}$ with contrary sign. Then, it is clear that

$$\mathrm{rank}(H_s) = \min\left\{N - 1, \frac{N_b}{2}\right\}, \tag{3.8}$$

where N_b is the number of nonzero elements in W^b. Due to the properties of matrix trace [17], we can get

$$\mathrm{rank}(S_s) = \mathrm{rank}(XH_s) \leq \min\{\mathrm{rank}(X), \mathrm{rank}(H_s)\}. \tag{3.9}$$



(a)

(b)

(c)

Figure 3: Recognition rate against the number of features used in the matching on the ORL database: (a) 3 training samples, (b) 4 training samples, and (c) 5 training samples.

Table 1: Comparison of recognition rates of Eigenface, Fisherface, DLDA, LPP, MFA, and DNDA on the ORL database.

Method	3 Training samples	4 Training samples	5 Training samples
Eigenface	86.64% (121)	91.65% (112)	94.05% (123)
Fisherface	87.46% (39)	91.42% (39)	93.35% (38)
DLDA	90.04% (38)	94.04% (39)	95.6% (37)
LPP	73.54% (91)	82.73% (98)	86.62% (99)
MFA	87.13% (27)	92% (39)	95.28% (41)
DNDA	91.07% (44)	94.69% (46)	96.12% (77)

Table 2: Comparison of recognition rates of Eigenface, Fisherface, DLDA, LPP, MFA, and DNDA on the Yale database.

Method	3 Training samples	4 Training samples	5 Training samples
Eigenfaces	76.79% (39)	80.14% (50)	82.39% (60)
Fisherfaces	80.96% (14)	84.27% (14)	90% (14)
DLDA	79.62% (12)	84.52% (11)	89.56% (14)
LPP	77.38% (44)	80.48% (59)	84.33% (59)
MFA	79.42% (26)	86.48% (23)	88.94% (24)
DNDA	82.42% (22)	88.62% (35)	90.61% (29)

In many FR cases, the number of pixels in a facial image is much larger than the number of available samples, that is, $d \gg N$. It tells us that the rank of S_s is at most $\min\{N-1, N_b/2\}$. Similarly, S_c can also be reformulated as

$$S_c = XH_cH_c^TX^T. \tag{3.10}$$

Here, $H_c \in R^{N\times N_w}$ is the intraclass index matrix consisting of all the N_w intraclass index vectors as columns, which is constructed according to the N_w nonzero elements in W^w. Similar to S_s, the rank of S_c is up to $\min\{N-1, N_w/2\}$. Based on the modified formulation, the optimal transformation matrix P can be obtained as

$$P^* = \arg\max_P \frac{P^TS_sP}{P^TS_cP} = \arg\max_P \frac{P^TG_sG_s^TP}{P^TG_cG_c^TP}, \quad G_s = XH_s,\ G_c = XH_c. \tag{3.11}$$

As the null space of S_s contributes little to classification, it is feasible to remove such subspace by projecting S_s into its range space. We apply the eigenvalue decomposition to S_s and unitize it through Turk's eigenanalysis method, while discarding those eigvectors whose corresponding eigvalues are zero, which do not take much power for discriminant analysis. Then, the discriminant information in S_c can be obtained by performing eigenanalysis on $\widetilde{S_c}$, which is gotten by transforming S_c into the range subspace of S_s. This algorithm can be implemented by the pseudocode shown in Algorithm 1.

Table 3: Comparison of recognition rates of Eigenface, Fisherface, DLDA, LPP, MFA, and DNDA on the UMIST database.

	Dimensionality	Recognition rate
Eigenfaces	99	89.84%
Fisherfaces	18	93.04%
DLDA	13	93.65%
LPP	93	92.95%
MFA	72	94.82%
DNDA	48	96.01%

DNDA has a computational complexity of $o(N_b^3)$ (N_b is the number of nonzero elements in W^b), as it preserves a similar procedure to DLDA ($o(c^3)$). Compared with Eigenface ($o(N^2 d)$) and Fisherface ($o(N^2 d)$), DNDA is still more efficient for feature extraction in high dimensionality if $d \gg N$.

4. Experiments

In this section, we investigate the performance of the proposed DNDA method for face recognition. Three popular face databases, ORL, Yale, and UMIST are used in the experiments. To verify the performance of DNDA, each experiment is compared with classical approaches: Eigenface [1], Fisherface [2], DLDA [6], LPP [11], and MFA [15]. The three nearest-neighbor classifier with Euclidean distance metric is employed to find the image in the database with the best match.

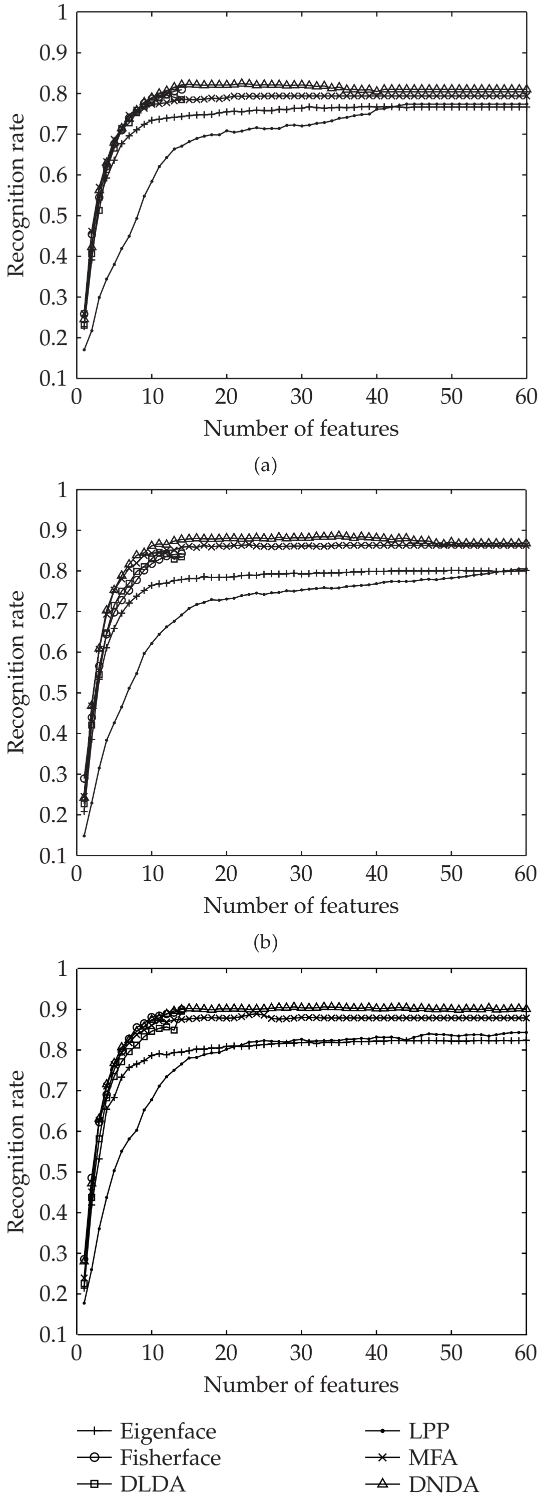
4.1. ORL database

In ORL database [18], there are 10 different images for each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling), and facial details (glasses/no glasses). All the images are taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The original images have size of 92 × 112 pixels with 256 gray levels; such one subject is illustrated in Figure 2(a).

The experiments are performed with different numbers of training samples. As there are 10 images for each subject, n ($n = 3, 4, 5$) of them are randomly selected for training and the remaining are used for testing. For each n, we perform 20 times to choose randomly the training set and the average recognition rate is calculated. Figure 3 illustrates the plot of recognition rate versus the number of features used in the matching for Eigenface, Fisherface, DLDA, LPP, MFA, and DNDA. The best performance obtained by each method and the corresponding dimension of reduced space in the bracket are shown in Table 1.

4.2. Yale database

The Yale Face Database [19] contains 165 grayscale images of 15 individuals. There are 11 images per subject, one per different lighting condition (left-light, center-light, right-light), facial expression (normal, happy, sad, sleepy, surprised, wink), and with/without glasses. Each images used in the experiments is 92 × 112 pixels with 256 gray levels. The facial images of one individual are illustrated in Figure 2(b).



(a)

(b)

(c)

Figure 4: Recognition rate against the number of features used in the matching on the Yale database: (a) 3 training samples, (b) 4 training samples and (c) 5 training samples.

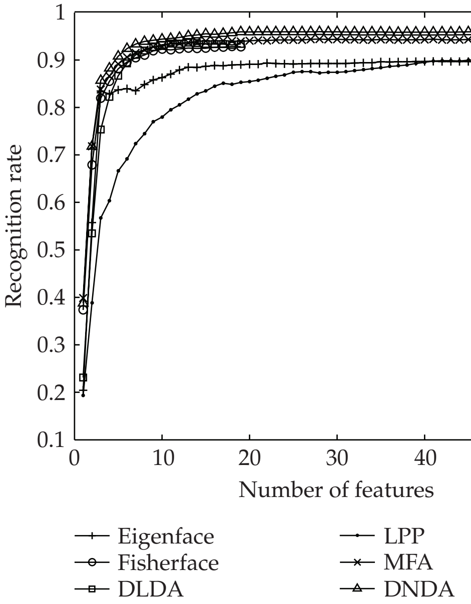


Figure 5: Recognition rate against the number of features used in the matching on the UMIST database.

The experimental implementation is the same as before. For each individual, $n\,(n = 3, 4, 5)$ images are randomly selected for training and the rest are used for testing. For each given n, we average the results over experiments repeated 20 times independently. Figure 4 illustrates the plot of recognition rate versus the number of features used in the matching for Eigenface, Fisherface, DLDA, LPP, MFA, and DNDA. The best results obtained in the experiments and the corresponding reduced dimension for each method is shown in Table 2.

4.3. UMIST database

The UMIST face database [20] consists of 564 images of 20 people. For simplicity, the Precropped version of the UMIST database is used in this experiment, where each subject covers a range of poses from profile to frontal views and a range of race/sex/appearance. The size of cropped image is 92 × 112 pixels with 256 gray levels. The facial images of one subject with different views are illustrated in Figure 2(c).

For each individual, we chose 8 images of different views distributed uniformly in the range 0–90° for training, and the rest are used for training. Figure 5 illustrates the plot of recognition rate versus the number of features used in the matching for Eigenface, Fisherface, DLDA, LPP, MFA, and DNDA. The best performance and the corresponding dimensionalities of the projected spaces for each method are shown in Table 3.

From the experiment results, it is very obvious that DNDA achieves higher accuracy than the other methods. This is probably due to the fact that DNDA is a two-stage local discriminant technique, different form LPP and MFA. Moreover, PCA is removed in DNDA preserving more discriminant information compared with others.

5. Conclusions

Inspired by DLDA, we propose in this paper a novel local discriminant feature extraction method called direct neighborhood discriminant analysis (DNDA). In order to avoid SSS problem, DNDA performs a two-stage eigenanalysis approach, which can be implemented efficiently by using Turk's method. Compared with other methods, PCA preprocessing is left out in DNDA with the immunity from the SSS problem. Experiments on ORL, Yale, and UMIST face databases show the effectiveness and robustness of our proposed method for face recognition. To get a better classification result, the improvement and extension of DNDA are to be taken into account in our future work.

Appendix

A. Proof of $2D - 2W = HH^T$

Given the graph weight matrix W with l nonzero elements, consider two matrices $M, N \in R^{N\times l}$. For each nonzero element in W, there is corresponding column in M and N with common location, respectively. Let $Z = \{(i,j) \mid W_{ij} \neq 0\}$ be the index set of nonzero elements in W. For the kth $(1 \leqslant k \leqslant l)$ nonzero element W_{ij} in W, the kth column of M, N is represented as

$$\begin{aligned} M_{(:,k)} &= \left[\overbrace{0 \ \cdots \ 0}^{i-1}, \overset{i}{1}, 0 \ \cdots \ \overset{N}{0}\right]^T, \\ N_{(:,k)} &= \left[\overbrace{0 \ \cdots \ 0}^{j-1}, \overset{j}{-1}, 0 \ \cdots \ \overset{N}{0}\right]^T. \end{aligned} \tag{A.1}$$

Then, it is easy to get

$$\begin{aligned} M_{(a,:)} M_{(b,:)}^T &= 0, \\ N_{(a,:)} N_{(b,:)}^T &= 0 \end{aligned} \tag{A.2}$$

for $a \neq b$ $(1 \leqslant a, b \leqslant N)$, and

$$\begin{aligned} M_{(a,:)} N_{(b,:)}^T &= 0 \quad \text{for } (a,b) \notin Z, \\ N_{(a,:)} M_{(b,:)}^T &= 0 \quad \text{for } (b,a) \notin Z, \end{aligned} \tag{A.3}$$

where $M_{(k,:)}$ and $N_{(k,:)}$ denote the kth row of M and N, respectively. Therefore, we can get

$$\begin{aligned} \left(MM^T\right)_{ij} &= \sum_{k=1}^{l} M_{ik} M_{jk} = \delta_{ij} \sum_{k=1}^{l} M_{ik} = \delta_{ij} \sum_{q=1}^{N} W_{iq}, \\ \left(NN^T\right)_{ij} &= \sum_{k=1}^{l} N_{ik} N_{jk} = \delta_{ij} \sum_{k=1}^{l} N_{ik} = \delta_{ij} \sum_{q=1}^{N} W_{qi}, \\ \left(MN^T\right)_{ij} &= \sum_{k=1}^{l} M_{ik} N_{jk} = W_{ij}, \\ \left(NM^T\right)_{ij} &= \sum_{k=1}^{l} N_{ik} M_{jk} = W_{ji}, \end{aligned} \tag{A.4}$$

where δ_{ij} is the Kronecker delta. Note that both matrix D and W are symmetry matrices, based on the above equations, it is easy to find out

$$\begin{aligned}(M-N)(M-N)^T &= MM^T + NN^T - MN^T - NM^T \\ &= D + D - W - W \\ &= 2D - 2W.\end{aligned} \tag{A.5}$$

It is easy to check that $H = M - N$, which completes the proof.

Acknowledgments

This work is supported in part by the Program for New Century Excellent Talents of Educational Ministry of China (NCET-06-0762) and Specialized Research Fund for the Doctoral Program of Higher Education (20060611009), and in part by Natural Science Foundations of Chongqing CSTC (CSTC2007BA2003 and CSTC2006BB2003), the National Natural Science Foundation of China under the Project Grant no. 60573125.

References

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[3] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized discriminant analysis for the small sample size problem in face recognition," *Pattern Recognition Letters*, vol. 24, no. 16, pp. 3079–3087, 2003.

[4] W.-S. Chen, P. C. Yuen, and J. Huang, "A new regularized linear discriminant analysis method to solve small sample size problems," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 19, no. 7, pp. 917–935, 2005.

[5] L.-F. Chen, H.-Y. M. Liao, M.-T. Ko, J.-C. Lin, and G.-J. Yu, "New LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713–1726, 2000.

[6] H. Yu and J. Yang, "A direct LDA algorithm for high dimensional data-with application to face recognition," *Pattern Recognition*, vol. 34, no. 10, pp. 2067–2070, 2001.

[7] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[8] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[9] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems 14 (NIPS '01)*, pp. 585–591, MIT Press, Cambridge, Mass, USA, 2002.

[10] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing Systems 16 (NIPS '03)*, pp. 153–160, MIT Press, Cambridge, Mass, USA, 2004.

[11] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacian faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.

[12] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, vol. 2, pp. 1208–1213, Beijing, China, October 2005.

[13] H.-T. Chen, H.-W. Chang, and T.-L. Liu, "Local discriminant embedding and its variants," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 2, pp. 846–853, San Diego, Calif, USA, June 2005.

[14] W. Zhang, X. Xue, H. Lu, and Y.-F. Guo, "Discriminant neighborhood embedding for classification," *Pattern Recognition*, vol. 39, no. 11, pp. 2240–2243, 2006.

[15] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.

[16] J. Ham, D. D. Lee, S. Mika, and B. Schölkopf, "A kernel view of the dimensionality reduction of manifolds," in *Proceedings of the 21th International Conference on Machine Learning (ICML '04)*, pp. 369–376, Banff, Alberta, Canada, July 2004.

[17] G. H. Golub and C. F. van Loan, *Matrix Computations*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, Md, USA, 3rd edition, 1996.

[18] Olivetti & Oracle Research Laboratory, The Olivetti & Oracle Research Laboratory Face Database of Faces, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[19] Yale Face Database, 2002, http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

[20] D. B. Graham and N. M. Allinson, "Characterizing virtual eigensignatures for general purpose face recognition," in *Face Recognition: From Theory to Applications*, H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie, and T. S. Huang, Eds., vol. 163 of *NATO ASI Series F, Computer and Systems Sciences*, pp. 446–456, Springer, Berlin, Germany, 1998.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 783278, 9 pages
doi:10.1155/2008/783278

Research Article

Intelligent Control of the Complex Technology Process Based on Adaptive Pattern Clustering and Feature Map

Wushan Cheng

Shanghai University of Engineering Science, Shanghai 200065, China

Correspondence should be addressed to Wushan Cheng, cwushan@163.com

Received 9 May 2008; Accepted 27 July 2008

Recommended by Carlo Cattani

A kind of fuzzy neural networks (FNNs) based on adaptive pattern clustering and feature map (APCFM) is proposed to improve the property of the large delay and time varying of the sintering process. By using the density clustering and learning vector quantization (LVQ), the sintering process is divided automatically into subclasses which have similar clustering center and labeled fitting number. Then these labeled subclass samples are taken into fuzzy neural network (FNN) to be trained; this network is used to solve the prediction problem of the burning through point (BTP). Using the 707 groups of actual training process data and the FNN to train APCFM algorithm, experiments prove that the system has stronger robustness and wide generality in clustering analysis and feature extraction.

1. Introduction

Sintering is the most widely used agglomeration process for iron ores and is a very important chain of iron making. In general, the process of sintering includes three major phases. First, it involves blending all the ores thoroughly according to certain proportions and adding water to the ore mix to produce particles. Second, the actual sintering operation is initiated by the ignition of the cokes as the raw mix passes under gas ignition. Finally, after traveling the length of the strand, the finished sinter is broken up, cooled, and screened [1, 2]. In the recent twenty years, many methods of integrity and fusion have been explored by the metallurgy and automation experts.

1.1. Mathematical model

According to the chemical/physical characteristics for sintering, a model was formulated as a series of differential equations to describe the relation between the thick martial, the ignition temperature, and the bellows temperature at the tail of the machine. For the time varying and randomness of the sintering process, many mechanisms have still not been understood. Although the dynamic model is tenable at a certain boundary condition, it is difficult to cover the whole process.

1.2. Neural network-based model

For the fast approach of neural network, a model can be established rapidly from the given input and output data, and it can also solve the problem of this long-time delay system. In general, genetic algorithm is used to optimize the parameters of the network and improve the generalization of the system, but it has still not been reported to be used in real-time control.

1.3. Rule-based model

The rule base, acknowledge, database, and inference machine can be constructed by the technology and operation experts' experience [3]. Rule base and inference machine are mainly used in estimating the process, analyzing cause, and deciding guideline. Acknowledge includes operation data, fact, mathematical model, and elicitation and unit knowledge. Database stores real-time data from production and equipment. Unfortunately, most results of this model are still simulation results.

2. Fuzzy neural network

In general, the dynamic behavior of a fuzzy logical controller is characterized by a set of linguistic control rules based on the knowledge of an expert [4].

Consider the fuzzy controller with Gaussian MFs and multiplication implication; the topology structure of fuzzy neural network is shown in Figure 1.

The fuzzy rule is as follows:

$$R^{(l)} : \text{if} \quad x_i = F_i^l,\ x_n = F_n^l, \tag{2.1}$$

then

$$y = G_i^l. \tag{2.2}$$

The input and output relationship is shown as

$$f(x) = \frac{\sum_{l=1}^{M} \overline{y}^{l} \left(\prod_{i=1}^{n} \mu_{F_i^l}(x_i) \right)}{\sum_{l=1}^{M} \left(\prod_{i=1}^{n} \mu_{F_i^l}(x_i) \right)}, \tag{2.3}$$

where $x = (x_1, x_2, \ldots, x_n)^T$ is the system input, M the is rule number, n is the input number, $\mu_{F_i^l}(x_i)$ is the membership function in the input $x_i = F_i^l$, and $\overline{y}^l$ is the value when the membership function equals the maximum in l rule. The fuzzy neural network [4] has five-layer structure.

The first layer is input variable layer. In this layer, the ith inputs are represented as x_i; the system can have n inputs.

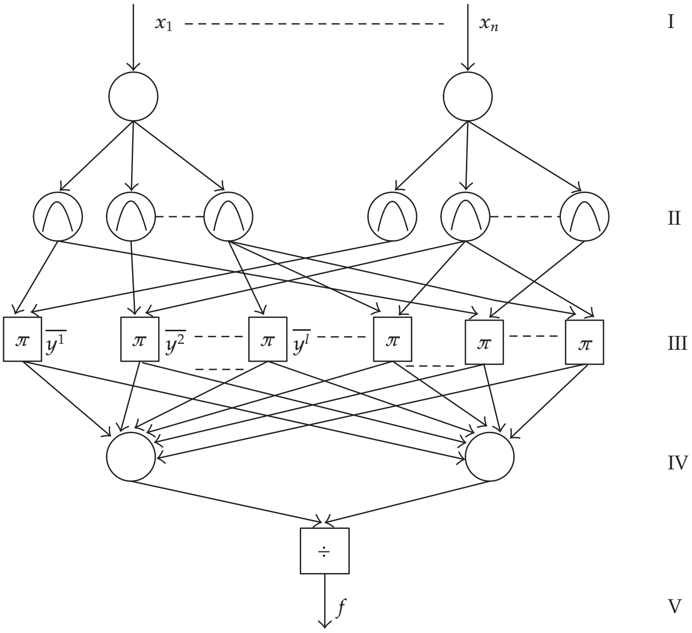


Figure 1: The topology structure of fuzzy neural network.

The second layer is membership layer. In this layer, each node performs the Gaussian function; the function is adopted as a membership function. The membership function of the input is defined as

$$\mu_{F_i^l}(x_i) = \exp\left[-\left(\frac{x_i - \overline{x}_i^l}{\sigma_i^l}\right)^2\right], \tag{2.4}$$

where $\overline{x}_i^l$ is the Gauss meaning of the rule input x_i, and σ_i^l is the square error.

The third layer is rule layer. The layer is used to implement the antecedent matching. The matching operation or the fuzzy and aggregation operation is chosen as the simple product operation. In this layer, summing is finished by neuron.

In addition to $\overline{y}^l$ between the third and the fouth layers, other layer weights equal 1.

The fifth layer is output of the fuzzy neural network.

Thus, the entire fuzzy neural network [5] needs to adjust $\overline{x}_i^l$, σ_i^l, $\overline{y}^l$ parameters to control the process. These parameters have specified signification; therefore, they are initialed by language information in order to improve learning convergence speed.

3. Adaptive pattern clustering and feature map network

3.1. Initial data space clustering

According to technology character and equipment requirement, density sintering speed and burning temperature are selected as input vectors; the temperature and pressure of 18 windboxes and the waste gas temperature are chosen as output vectors. The input space scatter diagram is obtained by using the input sample to do three-vector space map, and the

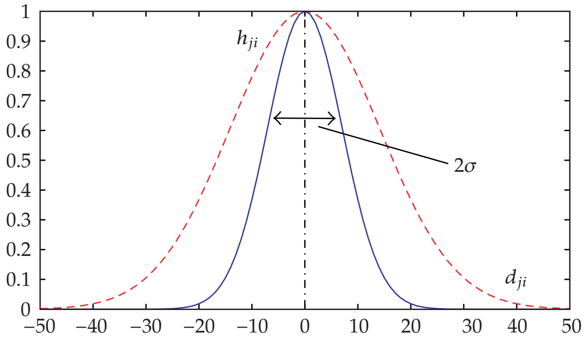


Figure 2: The trend of topology neighborhood coast line.

clustering center C_{ij} $(i = 1, 2, 3;\ j = 1, 2, \ldots, k_i)$ and subspace $[a_j, b_j]$ of every vector are found by utilizing feature extract based on density clustering. These rectangle areas are intersected with each other to form $k = k_1 \times k_2 \times k_3$ subregions.

3.2. Feature map

Feature map network developed by Kohonen is an unsupervised competitive learning cluster network in which only one neuron is on at any time. The map is an artificial system that emulates the brain in the visual system, and which includes three major phases [5–7].

Competitive phase: the inputs of the network can be written as vector by $X = [x_1, x_2, \ldots, x_m]^T$, and the synaptic weight vector of neuron j in the two-dimensional (2D) array is given by $w_j = [w_{j1}, w_{j2}, \ldots, w_{jm}]^T$, $j = 1, 2, \ldots, l$, where m is the local number of output neurons in the 2D array and l is the total number of the neurons of network. In order to find the best match of input vector x with the synaptic weight w_j, the multiplication $w_j^T x$ determined the center location of the exciting neuron's topology neighborhood and the maximum of $w_j^T x$ is equal to the Euclid norm in mathematics.

Cooperative phase: the winner neuron is located in the center of the cooperation neuron's topology neighborhood. We supposed that h_{ji} is the topology neighborhood whose center is the victory neuron i, and d_{ij} is the inclination distance between victory neuron i and excited neuron j. A classical selection of h_{ji} to satisfy these conditions is

$$h_{ji} = \exp\left(-\frac{d_{ji}^2}{2\sigma^2}\right), \tag{3.1}$$

where σ is the effective width of topology neighborhood. The trend of topology neighborhood is shown in Figure 2.

Self-adjusting phase: it includes self-ordering and converging stages; self-ordering formula is

$$\begin{aligned} w_j(n+1) &= w_j(n) + \eta(t)h_{j,i(x)}(n)(x(n) - w_j(n)), \\ \eta(n) &= \eta_0 \exp\left(-\frac{n}{\tau}\right). \end{aligned} \tag{3.2}$$

The equation is in converging stage; learning rate $\eta(n)$ is made smaller gradually.

3.3. Learning vector quantification

The learning vector quantification (LVQ) algorithm is used to adjust fine weight vector to improve quality in decision area by utilizing supervisor learning skill. The foundation method is first to find the average value of the attribute of every subclass on the basis of clustering, second to make a comparison between the average value of the subclass and the whole vector, and last to label the up-arrowhead tag with the larger values and the down-arrowhead with the smaller values. The set of every labeled subclasses may be expressed as the direction of its weight shifting. For this purpose, let the Lx_i stand for the tag of the input vector x_i, and let Lw_j stand for the tag of the weight w_j; the recursive function is defined as follows.

If $Lw_j = Lx_i$, then

$$w_j(n+1) = w_j(n) + \alpha_n(x_i - w_j(n)). \tag{3.3}$$

If $Lw_j \neq Lx_i$, then

$$w_j(n+1) = w_j(n) - \alpha_n(x_i - w_j(n)), \tag{3.4}$$

where $0 < \alpha_n < 1$.

Passing through a period of time iteratively, the subclasses with the same property may be converged together, and the other subclasses with different properties may be departed from each other.

In this paper, we use the actual data as the samples from sintering process. The input vectors are density, velocity, and ignition temperature, and the output vectors are the temperature and pressure of 18 windboxes and the temperature of waste gas.

4. Experiment

4.1. Analysis of the input in three-dimensional space

The distributing diagram of the two-year input samples in three-dimensional space is shown in Figure 3. We can obtain 12 subspaces by using the initialization clustering of the samples, which is based on the density of the samples, and maps the feature of 12 subspaces to form the topology structure, which is shown in Figure 4, and the center of every subspace is dotted in Figure 3.

4.2. Analysis of the input samples' classification

Computing the average value of every property for each subclass, respectively, such as density (D), velocity (V), and ignition temperature (T), and comparing the average value of the subclass property with the property of the whole samples, if the result of a subclass is bigger than the average value of the whole samples, we use up-arrowhead marking; otherwise we use down-arrowhead marking. The marking classification is listed in Table 1.

In this table, we can find 5 different large classes. Row 1 is a class, rows 2, 3, 5 are a class, rows 4, 7, 8, 10 are a class, rows 6, 9, 12 are a class, and row 11 is a class. Figure 5 shows the relations between the topology structure and the class table.

According to the characters of process and performance of equipments, we can get the property of each class in Figure 4.

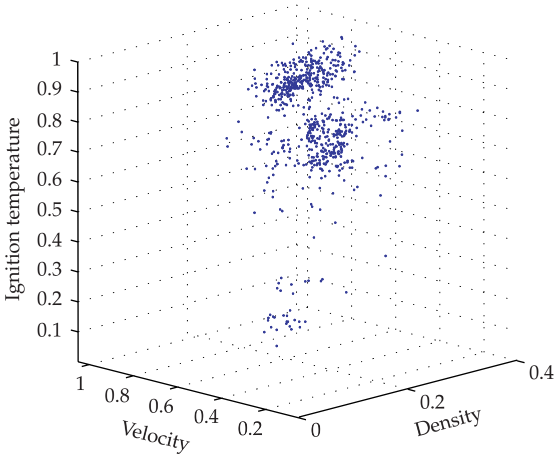


Figure 3: The distributing diagram of samples.

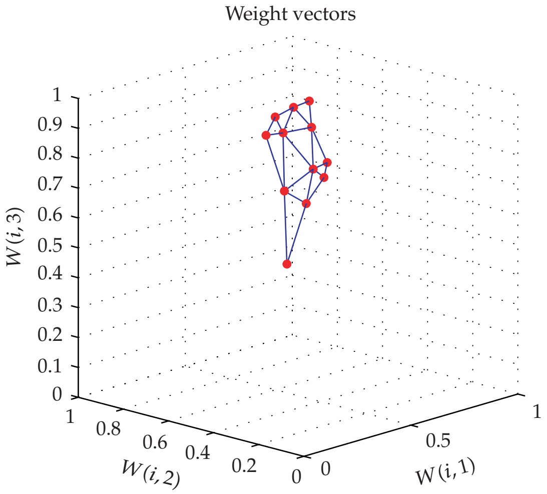


Figure 4: The topology structure of clustering.

Class 1 ($D \downarrow V \uparrow T \uparrow$). The samples of class denote the thick stuffing of sinter bed, high ignition temperature, and fast velocity, and it may cause raw ore.

Class 2 ($D \downarrow V \uparrow T \downarrow$). It denotes the thick stuffing of sinter bed, low ignition temperature, and fast velocity, and it causes easily raw ore, and the burning through point will be back to the strand tail.

Class 3 ($D \uparrow V \uparrow T \uparrow$). The class denotes loose stuffing on the sinter bed, high ignition temperature, and fast velocity, and it causes easily sintering for sintering process.

Class 4 ($D \downarrow V \downarrow T \downarrow$). The phenomenon shows the thick stuffing on the sinter bed, low ignition temperature, and slow velocity, and it is in accordance with the thick and slow sintering.

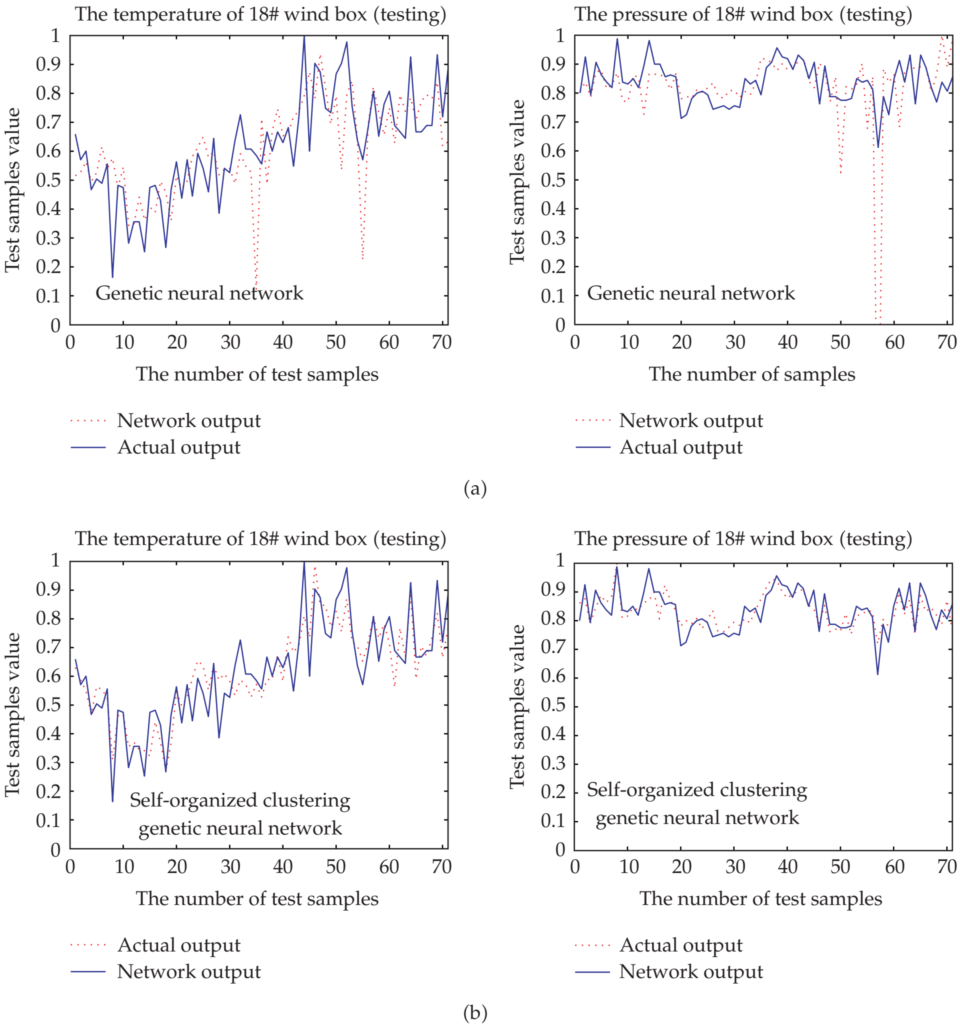


Figure 5: (a) The results of fuzzy neural network training (FNN). (b) The results of adaptive pattern clustering and feature map (APCFM).

Class 5 ($D \downarrow V \downarrow T \uparrow$). This state denotes the thick stuffing of sinter bed, high ignition temperature, and slow velocity, and we should direct our attention to the sinter bed earlier in order to avoid the oversintering.

4.3. Learning vector quantization

According to the analysis above, 12 subclasses have been readjusted into 5 classes. Now, retraining the whole input samples by using the LVQ network, the network is a characteristic studying of having teacher. The training network with the LVQ can improve the hitting accuracy of feature map that is proved by [6]. The network output can get the tag of the class when it enters the sample through the network. We show the step as follows. List the

Table 1: The setting of subclass property.

Subclass	Num	D	V	T	Compare results
1	65	0.6643	0.7986	0.8039	$D\downarrow V\uparrow T\uparrow$
2	29	0.6659	0.7893	0.6781	$D\downarrow V\uparrow T\downarrow$
3	40	0.6462	0.7208	0.3900	$D\downarrow V\uparrow T\downarrow$
4	92	0.7039	0.7779	0.8215	$D\uparrow V\uparrow T\uparrow$
5	17	0.6471	0.6775	0.6662	$D\downarrow V\uparrow T\downarrow$
6	32	0.6149	0.5921	0.5671	$D\downarrow V\downarrow T\downarrow$
7	95	0.7549	0.7505	0.8696	$D\uparrow V\uparrow T\uparrow$
8	27	0.6987	0.6771	0.8178	$D\uparrow V\uparrow T\uparrow$
9	96	0.6067	0.5184	0.7202	$D\downarrow V\downarrow T\downarrow$
10	75	0.7558	0.6892	0.8749	$D\uparrow V\uparrow T\uparrow$
11	34	0.6468	0.5361	0.7817	$D\downarrow V\downarrow T\uparrow$
12	105	0.5993	0.4557	0.7410	$D\downarrow V\downarrow T\downarrow$

input vectors P, the output vectors T, and the class of classificatory tag C:

$$\begin{aligned}
\mathrm{P} = [&0.75607, 0.81711, 0.78968, 0.6468, \ldots, 0.75626;\\
&0.67327, 0.66337, 0.64356, 0.5361, \ldots, 0.68317;\\
&0.96873, 0.92471, 0.95533, 0.7817, \ldots, 0.95406;],\\
\mathrm{T} = &\begin{bmatrix}1 & 2 & 4 & 5\cdots 3\end{bmatrix},\\
\mathrm{C} = &\begin{bmatrix}1 & 0 & 0 & 0 & \cdots & 0\\ 0 & 1 & 0 & 0 & \cdots & 0\\ 0 & 0 & 0 & 0 & \cdots & 1\\ 0 & 0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 0 & 1 & \cdots & 0\end{bmatrix}.
\end{aligned} \tag{4.1}$$

4.4. Training every subclass sample by using fuzzy neural network

The testing results are shown in Figures 5(a) and 5(b); Figure 5(a) is only genetic neural network testing results, and Figure 5(b) is the testing results by using the adaptive pattern clustering and feature map FNN. We compare the two figures and find out that FNN can obtain the trend of network output, but the precision is low. The adaptive pattern clustering and feature mapFNN can improve a high precision for network output and have a good generalization for the samples which belong to the same class.

5. Conclusion

In this paper, in order to predict the BTP, an APCFM reference and FNN system have been proposed to solve the challenging problem of the sinter production process, which is a typical nonlinear, time-varying, and multimode process, and is very difficult to solve using traditional methods. In our approach, a density clustering is used to determine the number of the initial input vectors consciously, and a feature map algorithm is used to extract data relevance property from different subclasses and improve the confidence of the vector. By using the teacher's instruction, LQV network can herd effectively feature categories together on this basis FNN algorithm. The constructed system has been trained with input sample consisting of 707 technology groups and measuring apparatus of two-year actual process and has obtained very good performance; especially, comparing APCFM+FNN with FNN [8, 9], the precision of training and testing has raised one time and three times, respectively, and the running time decreases more than one time, and it is satisfied with the demand of real time running and improving the robustness of the system.

Acknowledgment

This work is supported by the National Nature Science Foundation of China (Project no. 60274031).

References

[1] W. Cheng, "An application of adaptive genetic-neural algorithm to Sinter's BTP process," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, vol. 6, pp. 3356–3360, Shanghai, China, August 2004.

[2] W. Cheng, "Predictive fuzzy control applied to the mixing bunker," *Computer Application and Software*, vol. 21, no. 9, pp. 61–64, 2004.

[3] W. Cheng, "A building of the genetic-neural network for Sinter's burning through point," *Sintering and Pelletizing*, vol. 29, no. 5, pp. 18–22, 2004.

[4] M. J. Er, J. Liao, and J. Lin, "Fuzzy neural networks-based quality prediction system for sintering process," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 3, pp. 314–324, 2000.

[5] Q.-M. Feng, T. Li, X.-H. Fan, and T. Jiang, "Adaptive prediction system of sintering through point based on self-organize artificial neural network," *Transactions of Nonferrous Metals Society of China*, vol. 10, no. 6, pp. 804–807, 2000.

[6] M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*, Pearson Education, Upper Saddle River, NJ, USA, 2nd edition, 2002.

[7] Y. Wang, "Research on distributed hierarchical intelligent control for complex industrial system," *Computer Integrated Manufacturing System*, vol. 8, no. 7, pp. 551–554, 2002.

[8] B. Qin, M. Wu, and X. Wang, "Development and application of distributed integrated intelligent control system based on multi-agent system," *Mini-Micro Systems*, vol. 27, no. 7, pp. 1405–1408, 2006.

[9] J. Weijin and W. Pu, "Research on distributed solution and correspond consequence of complex system based on MAS," *Journal of Computer Research and Development*, vol. 43, no. 9, pp. 1615–1623, 2006.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 164808, 24 pages
doi:10.1155/2008/164808

Research Article

Shannon Wavelets Theory

Carlo Cattani

Department of Pharmaceutical Sciences (DiFarma), University of Salerno, Via Ponte don Melillo, Fisciano, 84084 Salerno, Italy

Correspondence should be addressed to Carlo Cattani, ccattani@unisa.it

Received 30 May 2008; Accepted 13 June 2008

Recommended by Cristian Toma

Shannon wavelets are studied together with their differential properties (known as connection coefficients). It is shown that the Shannon sampling theorem can be considered in a more general approach suitable for analyzing functions ranging in multifrequency bands. This generalization coincides with the Shannon wavelet reconstruction of $L_2(\mathbb{R})$ functions. The differential properties of Shannon wavelets are also studied through the connection coefficients. It is shown that Shannon wavelets are C^∞-functions and their any order derivatives can be analytically defined by some kind of a finite hypergeometric series. These coefficients make it possible to define the wavelet reconstruction of the derivatives of the C^ℓ-functions.

1. Introduction

Wavelets [1] are localized functions which are a very useful tool in many different applications: signal analysis, data compression, operator analysis, and PDE solving (see, e.g., [2] and references therein). The main feature of wavelets is their natural splitting of objects into different scale components [1, 3] according to the multiscale resolution analysis. For the $L_2(\mathbb{R})$ functions, that is, functions with decay to infinity, wavelets give the best approximation. When the function is localized in space, that is, the bottom length of the function is within a short interval (function with a compact support), such as pulses, any other reconstruction, but wavelets, leads towards undesirable problems such as the Gibbs phenomenon when the approximation is made in the Fourier basis. In this paper, it is shown that Shannon wavelets are the most expedient basis for the analysis of impulse functions (pulses) [4]. The approximation can be simply performed and the reconstruction by Shannon wavelets range in multifrequency bands. Comparing with the Shannon sampling theorem where the frequency band is only one, the reconstruction by Shannon wavelets can be done for functions ranging in different frequency bands. Shannon sampling theorem [5] plays a fundamental role in signal analysis and, in particular, for the reconstruction of a signal from a digital sampling. Under suitable hypotheses (on a given signal function) a few sets of values (samples) and a preliminary chosen basis (made by the sinc function) enable us to completely reconstruct the continuous signal. This reconstruction is alike the reconstruction of a function as a series expansion (such as polynomial, i.e., Taylor series, or trigonometric functions, i.e., Fourier series), but for the first time the reconstruction (in the sampling theorem) makes use of the sinc function, that is a localized function with decay to zero. Together with the Shannon sampling theorem (and reconstruction), also the wavelets series become very popular, as well as the bases with compact support. It has been recognized that on the sinc functions one can settle the family of Shannon wavelets. The main properties of these wavelets will be shown and discussed. Moreover, the connection coefficients [6–9] (also called refinable integrals) will be computed by giving some finite formulas for any order derivatives (see also some preliminary results in [2, 10–12]). These coefficients enable us to define any order derivatives of the Shannon scaling and wavelet basis and it is shown that also the derivatives are orthogonal.

2. Shannon Wavelets

Sinc function or Shannon scaling function is the starting point for the definition of the Shannon wavelet family [11]. It can be shown that the Shannon wavelets coincide with the real part of the harmonic wavelets [2, 10, 13, 14], which are the band-limited complex functions

$$\psi_k^n(x) \overset{\text{def}}{\equiv} 2^{n/2} \frac{e^{4\pi i(2^n x-k)} - e^{2\pi i(2^n x-k)}}{2\pi i(2^n x - k)}, \tag{2.1}$$

with $n, k \in \mathbb{Z}$. Harmonic wavelets form an orthonormal basis and give rise to a multiresolution analysis [1–3, 14, 15]. In the frequency domain, they are very well localized and defined on compact support intervals, but they have a very slow decay in the space variable. However, in dealing with real problems it is more expedient to make use of real basis. By focussing on the real part of the harmonic family, we can take advantage of the basic properties of harmonic wavelets together with a more direct physical interpretation of the basis.

Let us take, as scaling function $\varphi(x)$, the sinc function (Figure 1)

$$\varphi(x) = \operatorname{sinc} x \overset{\text{def}}{=} \frac{\sin \pi x}{\pi x} = \frac{e^{\pi i x} - e^{-\pi i x}}{2\pi i x} \tag{2.2}$$

and for the dilated and translated instances

$$\begin{aligned} \varphi_k^n(x) &= 2^{n/2}\varphi(2^n x - k) = 2^{n/2}\frac{\sin \pi(2^n x - k)}{\pi(2^n x - k)} \\ &= 2^{n/2}\frac{e^{\pi i(2^n x-k)} - e^{-\pi i(2^n x-k)}}{2\pi i(2^n x - k)}. \end{aligned} \tag{2.3}$$

The parameters n, k give, respectively, a compression (dilation) of the basic function (2.2) and a translation along the x-axis. The family of translated instances $\{\varphi(x-k)\}$ is an orthonormal basis for the banded frequency functions [5] (Shannon theorem). For this reason, they can be used to define the Shannon multiresolution analysis as follows. The scaling functions do not represent a basis, in a functional space, therefore we need to define a family of

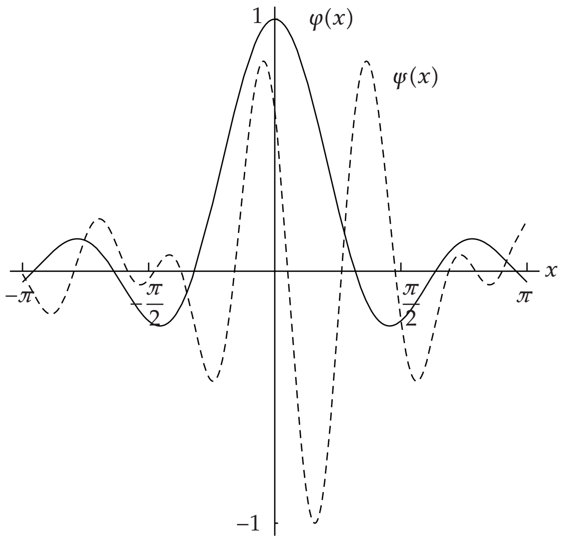


Figure 1: Shannon scaling function $\varphi(x)$ (thick line) and wavelet (dashed line) $\psi(x)$.

functions (based on scaling) which are a basis; they are called the wavelet functions and the corresponding analysis the multiresolution analysis.

Let

$$\widehat{f}(\omega) = \widehat{f(x)} \overset{\text{def}}{=} \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{-i\omega x}dx \tag{2.4}$$

be the Fourier transform of the function $f(x) \in L_2(\mathbb{R})$ and

$$f(x) = 2\pi\int_{-\infty}^{\infty} \widehat{f}(\omega)e^{i\omega x}\mathrm{d}\omega \tag{2.5}$$

its inverse transform. The Fourier transform of (2.2) gives us

$$\widehat{\varphi}(\omega) = \frac{1}{2\pi}\chi(\omega + 3\pi) = \begin{cases} \dfrac{1}{(2\pi)}, & -\pi \le \omega < \pi, \\ 0, & \text{elsewhere,} \end{cases} \tag{2.6}$$

with

$$\chi(\omega) = \begin{cases} 1, & 2\pi \le \omega < 4\pi, \\ 0, & \text{elsewhere}\,. \end{cases} \tag{2.7}$$

Analogously for the dilated and translated instances of scaling function it is

$$\widehat{\varphi}_k^n(\omega) = \frac{2^{-n/2}}{2\pi}e^{-i\omega(k+1)/2^n}\chi\left(\frac{\omega}{2^n} + 3\pi\right). \tag{2.8}$$

From the given scaling function, it is possible to define the corresponding wavelet function [1, 15] according to the following.

Theorem 2.1. *The Shannon wavelet, in the Fourier domain, is*

$$\widehat{\psi}(\omega) = \frac{1}{2\pi} e^{-i\omega}[\chi(2\omega) + \chi(-2\omega)]. \tag{2.9}$$

Proof. It can be easily shown that the scaling function (2.6) fulfills the condition

$$\widehat{\varphi}(\omega) = H\left(\frac{\omega}{2}\right)\widehat{\varphi}\left(\frac{\omega}{2}\right), \tag{2.10}$$

which characterizes the multiresolution analysis [1] with

$$H\left(\frac{\omega}{2}\right) = \chi(\omega + 3\pi). \tag{2.11}$$

Thus the corresponding wavelet function can be defined as [1, 15]

$$\widehat{\psi}(\omega) = e^{-i\omega}\overline{H\left(\frac{\omega}{2} \pm 2\pi\right)}\widehat{\varphi}\left(\frac{\omega}{2}\right). \tag{2.12}$$

With $H(\omega/2 - 2\pi)$ we have

$$\begin{aligned}
\widehat{\psi}(\omega) &= e^{-i\omega}\overline{H\left(\frac{\omega}{2} - 2\pi\right)}\widehat{\varphi}\left(\frac{\omega}{2}\right) \\
&= e^{-i\omega}\chi(\omega + 3\pi - 2\pi)\frac{1}{2\pi}\chi\left(\frac{\omega}{2} + 3\pi\right) \\
&= \frac{1}{2\pi}e^{-i\omega}\chi(\omega + \pi)\chi\left(\frac{\omega}{2} + 3\pi\right) \\
&= \frac{1}{2\pi}e^{-i\omega}\chi(2\omega),
\end{aligned} \tag{2.13}$$

then analogously with $H(\omega/2 + 2\pi)$ we obtain

$$\widehat{\psi}(\omega) = \frac{1}{2\pi}e^{-i\omega}\chi(-2\omega), \tag{2.14}$$

so that (2.9) follows. □

For the whole family of dilated-translated instances, it is

$$\widehat{\psi}^n_k(\omega) = \frac{2^{-n/2}}{2\pi}e^{-i\omega(k+1)/2^n}\left[\chi\left(\frac{\omega}{2^{n-1}}\right) + \chi\left(\frac{-\omega}{2^{n-1}}\right)\right]. \tag{2.15}$$

The Shannon wavelet function in the real domain can be obtained from (2.9) by the inverse Fourier transform (Figure 1)

$$\begin{aligned}
\psi(x) &= \frac{\sin\pi(x - 1/2) - \sin 2\pi(x - 1/2)}{\pi(x - 1/2)} \\
&= \frac{e^{-2i\pi x}(-i + e^{i\pi x} + e^{3i\pi x} + ie^{4i\pi x})}{(\pi - 2\pi x)},
\end{aligned} \tag{2.16}$$

and by the space shift and compression we have the whole family of dilated and translated instances:

$$\psi^n_k(x) = 2^{n/2}\frac{\sin\pi(2^n x - k - 1/2) - \sin 2\pi(2^n x - k - 1/2)}{\pi(2^n x - k - 1/2)}. \tag{2.17}$$

By summarizing (2.3) and (2.17), the Shannon wavelet theory is based on the following functions [11]:

$$\varphi_k^n(x) = 2^{n/2}\frac{\sin \pi(2^n x - k)}{\pi(2^n x - k)},$$
$$\psi_k^n(x) = 2^{n/2}\frac{\sin \pi(2^n x - k - 1/2) - \sin 2\pi(2^n x - k - 1/2)}{\pi(2^n x - k - 1/2)} \tag{2.18}$$

in the space domain, and collecting (2.8) and (2.15), we have in the frequency domain

$$\widehat{\varphi}_k^n(\omega) = \frac{2^{-n/2}}{2\pi}e^{-i\omega k/2^n}\chi\left(\frac{\omega}{2^n}+3\pi\right),$$
$$\widehat{\psi}_k^n(\omega) = -\frac{2^{-n/2}}{2\pi}e^{-i\omega(k+1/2)/2^n}\left[\chi\left(\frac{\omega}{2^{n-1}}\right)+\chi\left(\frac{-\omega}{2^{n-1}}\right)\right]. \tag{2.19}$$

The inner product is defined as

$$\langle f,g\rangle \overset{\text{def}}{\equiv} \int_{-\infty}^{\infty} f(x)\overline{g(x)}\mathrm{d}x, \tag{2.20}$$

which, according to the Parseval equality, can be expressed also as

$$\langle f,g\rangle \overset{\text{def}}{\equiv} \int_{-\infty}^{\infty} f(x)\overline{g(x)}\mathrm{d}x = 2\pi\int_{-\infty}^{\infty}\widehat{f}(\omega)\overline{\widehat{g}(\omega)}\mathrm{d}\omega = 2\pi\langle \widehat{f},\widehat{g}\rangle, \tag{2.21}$$

where the bar stands for the complex conjugate.

With respect to the inner product (2.20), we can show the following theorem [11].

Theorem 2.2. *Shannon wavelets are orthonormal functions in the sense that*

$$\langle \psi_k^n(x),\psi_h^m(x)\rangle = \delta^{nm}\delta_{hk}, \tag{2.22}$$

with δ^{nm}, δ_{hk} being the Kroenecker symbols.

Proof.

$$\begin{aligned}
&\langle \psi_k^n(x),\psi_h^m(x)\rangle \\
&= 2\pi\langle \widehat{\psi}_k^n(\omega),\widehat{\psi}_h^m(\omega)\rangle \\
&= 2\pi\int_{-\infty}^{\infty}\frac{2^{-n/2}}{2\pi}e^{-i\omega(k+1/2)/2^n}\left[\chi\left(\frac{\omega}{2^{n-1}}\right)+\chi\left(\frac{-\omega}{2^{n-1}}\right)\right]\frac{2^{-m/2}}{2\pi}e^{i\omega(h+1/2)/2^m}\left[\chi\left(\frac{\omega}{2^{m-1}}\right)+\chi\left(\frac{-\omega}{2^{m-1}}\right)\right]\mathrm{d}\omega \\
&= \frac{2^{-(n+m)/2}}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega(k+1/2)/2^n+i\omega(h+1/2)/2^m}\left[\chi\left(\frac{\omega}{2^{n-1}}\right)+\chi\left(\frac{-\omega}{2^{n-1}}\right)\right]\left[\chi\left(\frac{\omega}{2^{m-1}}\right)+\chi\left(\frac{-\omega}{2^{m-1}}\right)\right]\mathrm{d}\omega
\end{aligned} \tag{2.23}$$

which is zero for $n \neq m$. For $n = m$ it is

$$\langle \psi_k^n(x),\psi_h^n(x)\rangle = \frac{2^{-n}}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega(h-k)/2^n}\left[\chi\left(\frac{\omega}{2^{n-1}}\right)+\chi\left(\frac{-\omega}{2^{n-1}}\right)\right]\mathrm{d}\omega \tag{2.24}$$

and, according to (2.7), by the change of variable $\xi = \omega/2^{n-1}$

$$\langle \varphi^n_k(x), \varphi^n_h(x)\rangle = \frac{1}{4\pi}\left[\int_{-4\pi}^{-2\pi} e^{-2i(h-k)\xi}\mathrm{d}\xi + \int_{2\pi}^{4\pi} e^{-2i(h-k)\xi}\mathrm{d}\xi\right]. \tag{2.25}$$

For $h = k$ (and $n = m$), it is trivially

$$\langle \varphi^n_k(x), \varphi^n_k(x)\rangle = 1. \tag{2.26}$$

For $h \neq k$, it is

$$\int_{2\pi}^{4\pi} e^{-2i(h-k)\xi}\mathrm{d}\xi = \frac{i}{2(h-k)}\left(e^{-4i\pi(h-k)} - e^{-8i\pi(h-k)}\right) = 0\,, \tag{2.27}$$

and analogously $\int_{-4\pi}^{-2\pi} e^{-2i(h-k)\xi}\mathrm{d}\xi = 0$. □

Moreover, we have the following theorem [11].

Theorem 2.3. *The translated instances of the Shannon scaling functions $\varphi^n_k(x)$, at the level $n = 0$, are orthogonal in the sense that*

$$\langle \varphi^0_k(x), \varphi^0_h(x)\rangle = \delta_{kh}\,, \tag{2.28}$$

being $\varphi^0_k(x) \overset{\text{def}}{=} \varphi(x-k)$.

Proof. It is

$$\begin{aligned}\langle \varphi^n_k(x), \varphi^m_h(x)\rangle &= 2\pi\langle \widehat{\varphi}^n_k(\omega), \widehat{\varphi}^m_h(\omega)\rangle \\ &= 2\pi\int_{-\infty}^{\infty} \frac{2^{-n/2}}{2\pi} e^{-i\omega k/2^n}\chi\left(\frac{\omega}{2^n} + 3\pi\right)\frac{2^{-m/2}}{2\pi} e^{i\omega h/2^m}\chi\left(\frac{\omega}{2^m} + 3\pi\right)\mathrm{d}\omega \\ &= \frac{2^{-(n+m)/2}}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega(k/2^n - h/2^m)}\chi\left(\frac{\omega}{2^n} + 3\pi\right)\chi\left(\frac{\omega}{2^m} + 3\pi\right)\mathrm{d}\omega\,.\end{aligned} \tag{2.29}$$

When $m = n$, we have

$$\langle \varphi^n_k(x), \varphi^n_h(x)\rangle = \frac{2^{-n}}{2\pi}\int_{-2^n\pi}^{2^n\pi} e^{-i\omega(k-h)/2^n}\mathrm{d}\omega = 2^n\frac{\sin[(h-k)\pi]}{(h-k)\pi}\,. \tag{2.30}$$

Since $h, k \in \mathbb{Z}$, there follows that

$$\frac{\sin[(h-k)\pi]}{(h-k)\pi} = \left\{\begin{matrix} 1, & h = k \\ 0, & h \neq k \end{matrix}\right\} = \delta_{kh}, \tag{2.31}$$

that is,

$$\langle \varphi^n_k(x), \varphi^n_h(x)\rangle = \delta_{kh}. \tag{2.32}$$

When $m \neq n$, let's say $m < n$, we have

$$\langle \varphi^n_k(x), \varphi^m_h(x)\rangle = \frac{2^{-(n+m)/2}}{2\pi}\int_{-2^m\pi}^{2^m\pi} e^{-i\omega(k/2^n - h/2^m)}\mathrm{d}\omega, \tag{2.33}$$

that is,

$$\langle \varphi^n_k(x), \varphi^m_h(x)\rangle = 2^{(m+n)/2}\frac{\sin[(h-2^{m-n}k)\pi]}{(h-2^{m-n}k)\pi}. \tag{2.34}$$

When $m \neq n$, the last expression is always different from zero, in fact (since $m < n$)

$$\sin\left[\left(h - \frac{k}{2^{|m-n|}}\right)\pi\right] = 0 \Longrightarrow \left[h - \frac{k}{2^{|m-n|}}\right]\pi = s\pi\,, \quad s \in \mathbb{Z} \tag{2.35}$$

that is,

$$h = s + \frac{k}{2^{|m-n|}}, \quad h, k, s \in \mathbb{Z} \tag{2.36}$$

and $h \in \mathbb{Z}$ only if $m = n$. Therefore, in order to have the orthogonality it must be$m = n$, so that

$$\langle \varphi_k^n(x), \varphi_h^n(x) \rangle = 2^n \delta_{kh}. \tag{2.37}$$

and, in particular, when $n = 0$,

$$\langle \varphi_k^0(x), \varphi_h^0(x) \rangle = \delta_{kh}. \tag{2.38}$$

□

As a consequence of this proof we have that

$$\varphi_k^0(h) = \delta_{kh} \quad (h, k \in \mathbb{Z}). \tag{2.39}$$

The scalar product of the (Shannon) scaling functions with the corresponding wavelets is characterized by the following [11].

Theorem 2.4. *The translated instances of the Shannon scaling functions $\varphi_k^n(x)$, at the level $n = 0$, are orthogonal to the Shannon wavelets in the sense that*

$$\langle \varphi_k^0(x), \psi_h^m(x) \rangle = 0, \quad m \geq 0, \tag{2.40}$$

being $\varphi_k^0(x) \overset{\text{def}}{=} \varphi(x - k)$.

Proof. It is

$$\begin{aligned}
&\langle \varphi_k^n(x), \psi_h^m(x) \rangle \\
&\quad = 2\pi \langle \widehat{\varphi}_k^n(\omega), \widehat{\psi}_h^m(\omega) \rangle \\
&\quad = 2\pi \int_{-\infty}^{\infty} 2^{-n/2} e^{-i\omega k/2^n} \chi\left(\frac{\omega}{2^n} + 3\pi\right) \frac{2^{-m/2}}{2\pi} e^{i\omega(h+1/2)/2^m} \left[\chi\left(\frac{\omega}{2^{m-1}}\right) + \chi\left(\frac{-\omega}{2^{m-1}}\right)\right] \mathrm{d}\omega \\
&\quad = 2^{-(n+m)/2} \int_{-\infty}^{\infty} e^{-i\omega k/2^n + i\omega(h+1/2)/2^m} \chi\left(\frac{\omega}{2^n} + 3\pi\right) \left[\chi\left(\frac{\omega}{2^{m-1}}\right) + \chi\left(\frac{-\omega}{2^{m-1}}\right)\right] \mathrm{d}\omega
\end{aligned} \tag{2.41}$$

which is zero for $m \geq n \geq 0$ (since, according to (2.7), the compact support of the characteristic functions do not intersect).

On the contrary, it can be easily seen that, for $m < n$, it is

$$\begin{aligned}
\langle \varphi_k^n(x), \psi_h^m(x) \rangle &= 2^{-(n+m)/2} \int_{2^m\pi}^{2^n\pi} e^{-i\omega k/2^n + i\omega(h+1/2)/2^m} \mathrm{d}\omega \\
&= -\frac{2^{1+(m+n)/2}\left(ie^{i\pi[2^{-m+n-1}(1+2h)-k]} + e^{i\pi(h-2^{m-n}k)}\right)}{2^n(1+2h) - 2^{1+m}k}
\end{aligned} \tag{2.42}$$

and this product, in general, does not vanish. □

3. Reconstruction of a Function by Shannon Wavelets

Let $f(x) \in L_2(\mathbb{R})$ be a function such that for any value of the parameters $n, k \in \mathbb{Z}$, it is

$$\left|\int_{-\infty}^{\infty} f(x)\varphi_k^0(x)\mathrm{d}x\right| \le A_k^n < \infty, \qquad \left|\int_0^{\infty} f(x)\psi_k^n(x)\mathrm{d}x\right| \le B_k^n < \infty, \tag{3.1}$$

and $B \subset L_2(\mathbb{R})$ the Paley-Wiener space, that is, the space of band-limited functions such that,

$$\operatorname{supp} \widehat{f} \subset [-b, b], \quad b < \infty. \tag{3.2}$$

For the representation with respect to the basis (2.18), it is $b = \pi$. According to the sampling theorem (see, e.g., [5]) we have the the following.

Theorem 3.1 (Shannon). *If* $f(x) \in L_2(\mathbb{R})$ *and* $\operatorname{supp} \widehat{f} \subset [-\pi, \pi]$, *the series*

$$f(x) = \sum_{k=-\infty}^{\infty} \alpha_k \varphi_k^0(x) \tag{3.3}$$

uniformly converges to $f(x)$, *and*

$$\alpha_k = f(k). \tag{3.4}$$

Proof. In order to compute the values of the coefficients, we have to evaluate the series in correspondence of the integer:

$$f(h) = \sum_{k=-\infty}^{\infty} \alpha_k \varphi_k^0(h) = \sum_{k=-\infty}^{\infty} \alpha_k \delta_{kh} = \alpha_h, \tag{3.5}$$

having taken into account (2.39).

The convergence follows from the hypotheses on $f(x)$. In particular, the importance of the band-limited frequency can be easily seen by applying the Fourier transform to (3.3):

$$\begin{aligned} \widehat{f}(\omega) &= \sum_{k=-\infty}^{\infty} f(k)\widehat{\varphi}_k^0(x) \\ &\overset{(2.8)}{=} \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} f(k)e^{-i\omega k}\chi(\omega + 3\pi) \\ &= \frac{1}{2\pi}\chi(\omega + 3\pi) \sum_{k=-\infty}^{\infty} f(k)e^{-i\omega k} \end{aligned} \tag{3.6}$$

so that

$$\widehat{f}(\omega) = \begin{cases} \dfrac{1}{2\pi} \displaystyle\sum_{k=-\infty}^{\infty} f(k)e^{-i\omega k}, & \omega \in [-\pi, \pi], \\ 0, & \omega \notin [-\pi, \pi]. \end{cases} \tag{3.7}$$

In other words, if the function is band limited (i.e., with compact support in the frequency domain), it can be completely reconstructed by a discrete Fourier series. The Fourier coefficients are the values of the function $f(x)$ sampled at the integers. □

As a generalization of the Paley-Wiener space, and in order to generalize the Shannon theorem, we define the space $\mathcal{B}_{\varphi} \supseteq \mathcal{B}$ of functions $f(x)$ such that the integrals

$$\begin{aligned} \alpha_k &= \langle f(x), \varphi_k^0(x) \rangle = \int_{-\infty}^{\infty} f(x) \varphi_k^0(x) \mathrm{d}x, \\ \beta_k^n &= \langle f(x), \psi_k^n(x) \rangle = \int_{-\infty}^{\infty} f(x) \psi_k^n(x) \mathrm{d}x \end{aligned} \tag{3.8}$$

exist and are finite. According to (2.20) and (2.21), it is in the Fourier domain that

$$\begin{aligned} \alpha_k &= 2\pi \langle \widehat{f(x)}, \widehat{\varphi_k^0(x)} \rangle = \int_{-\infty}^{\infty} \widehat{f}(\omega) \widehat{\varphi}_k^0(\omega) \mathrm{d}\omega = \int_0^{2\pi} \widehat{f}(\omega) e^{i\omega k} \mathrm{d}\omega, \\ \beta_k^n &= 2\pi \langle \widehat{f(x)}, \widehat{\psi_k^n(x)} \rangle = \cdots = 2^{-n/2} \int_{2^{n+1}\pi}^{2^{n+2}\pi} \widehat{f}(\omega) e^{i\omega k/2^n} \mathrm{d}\omega . \end{aligned} \tag{3.9}$$

Let us prove the following.

Theorem 3.2 (Shannon generalization). *If $f(x) \in B_{\varphi} \subset L_2(\mathbb{R})$ and* $\operatorname{supp} \widehat{f} \subseteq \mathbb{R}$, *the series*

$$f(x) = \sum_{h=-\infty}^{\infty} \alpha_h \varphi_h^0(x) + \sum_{n=0}^{\infty} \sum_{k=-\infty}^{\infty} \beta_k^n \psi_k^n(x) \tag{3.10}$$

converges to $f(x)$, with α_h and β_k^n given by (3.8) *and* (3.9). *In particular, when* $\operatorname{supp} \widehat{f} \subseteq [-2^N\pi, 2^N\pi]$, *it is*

$$f(x) = \sum_{h=-\infty}^{\infty} \alpha_h \varphi_h^0(x) + \sum_{n=0}^{N} \sum_{k=-\infty}^{\infty} \beta_k^n \psi_k^n(x). \tag{3.11}$$

Proof. The representation (3.10) follows from the orthogonality of the scaling and Shannon wavelets (Theorems 2.2, 2.3, 2.4). The coefficients, which exist and are finite, are given by (3.8). The convergence of the series is a consequence of the wavelet axioms.

It should be noticed that

$$\operatorname{supp} \widehat{f} = [-\pi, \pi] \bigcup_{n=0,\ldots,\infty} [-2^{n+1}\pi, \ -2^n\pi] \cup [2^n\pi, 2^{n+1}\pi] \tag{3.12}$$

so that for a band-limited frequency signal, that is, for a signal whose frequency belongs to the first band $[-\pi, \pi]$, this theorem reduces to the Shannon. But, more in general, one has to deal with a signal whose frequency range in different bands, even if practically banded, since it is $N < \infty$. In this case, we have some nontrivial contributions to the series coefficients from all the bands, ranging from $[-2^N\pi, 2^N\pi]$:

$$\operatorname{supp} \widehat{f} = [-\pi, \pi] \bigcup_{n=0,\ldots,N} [-2^{n+1}\pi, \ -2^n\pi] \cup [2^n\pi, 2^{n+1}\pi]. \tag{3.13}$$

In the frequency domain, (3.10) gives

$$\begin{aligned}
f(x) &= \sum_{h=-\infty}^{\infty} \alpha_h \varphi_h^0(x) + \sum_{n=0}^{\infty}\sum_{k=-\infty}^{\infty} \beta_k^n \psi_k^n(x),\\
\widehat{f}(\omega) &= \sum_{h=-\infty}^{\infty} \alpha_h \widehat{\varphi}_h^0(\omega) + \sum_{n=0}^{\infty}\sum_{k=-\infty}^{\infty} \beta_k^n \widehat{\psi}_k^n(\omega),\\
\widehat{f}(\omega) &\overset{(2.19)}{=} \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} \alpha_h e^{-i\omega h}\chi(\omega+3\pi) + \frac{1}{2\pi}\sum_{n=0}^{\infty}\sum_{k=-\infty}^{\infty} 2^{-n/2}\beta_k^n e^{-i\omega(k+1)/2^n}\left[\chi\left(\frac{\omega}{2^{n-1}}\right)+\chi\left(\frac{-\omega}{2^{n-1}}\right)\right],
\end{aligned} \tag{3.14}$$

that is,

$$\begin{aligned}
\widehat{f}(\omega) = {} & \frac{1}{2\pi}\chi(\omega+3\pi)\sum_{h=-\infty}^{\infty} \alpha_h e^{-i\omega h}\\
&+ \frac{1}{2\pi}\chi\left(\frac{\omega}{2^{n-1}}\right)\sum_{n=0}^{\infty}\sum_{k=-\infty}^{\infty} 2^{-n/2}\beta_k^n e^{-i\omega(k+1)/2^n}\\
&+ \frac{1}{2\pi}\chi\left(\frac{-\omega}{2^{n-1}}\right)\sum_{n=0}^{\infty}\sum_{k=-\infty}^{\infty} 2^{-n/2}\beta_k^n e^{-i\omega(k+1)/2^n}.
\end{aligned} \tag{3.15}$$

There follows that the Fourier transform is made by the composition of coefficients at different frequency bands. When $\beta_k^n = 0$, for all $n, k \in \mathbb{Z}$, we obtain the Shannon sampling theorem as a special case. □

Of course, if we limit the dilation factor $n \leq N < \infty$, for a truncated series, we have the approximation of $f(x)$, given by

$$f(x) \cong \sum_{h=-S}^{S} \alpha_h \varphi(x-h) + \sum_{n=0}^{N}\sum_{k=-M}^{M} \beta_k^n \psi_k^n(x). \tag{3.16}$$

By rearranging the many terms of the series with respect to the different scales, for a fixed N we have

$$\begin{aligned}
f(x) &\cong \sum_{h=-S}^{S} \alpha_h \varphi(x-h) + \sum_{n=0}^{N} f_n(x),\\
f_n(x) &= \sum_{k=-M}^{M} \beta_k^n \psi_k^n(x),
\end{aligned} \tag{3.17}$$

where $f_n(x)$ represent the component of the function $f(x)$ at the scale $0 \leq n \leq N$ (i.e., in the band $[2^n\pi, 2^{n+1}\pi]$), and $f(x)$ results from a multiscale approximation or better from the multiband reconstruction.

3.1. Examples

Let us first compute the approximate wavelet representation of the even function

$$f(x) = e^{-4x^2} \cos 2\pi x. \tag{3.18}$$

The bottom length (i.e., the main part) of the function $f(x)$ is concentrated in the interval $[-0.2, 0.2]$. With a low scale $n = 3$, we can have a good approximation (Figures 2, 4) of the function even with a small number k of translation. In fact, with $|k| \leq 3$ the absolute value of the approximation error is less than 7% (see Figure 4). The higher number of the translation parameter k improves the approximation of the function on its "tails," in the sense that by increasing the number of translation parameters k the oscillation on "tails" is reduced. We can see that with $|k| \leq 10$ the approximation error is reduced up to 3%. Moreover, the approximation error tends to zero with $|x| \to \infty$.

The multiscale representation is given by

$$\begin{gathered} f(x) \cong \alpha_0 \varphi(x) + \sum_{n=0}^{3} f_n(x), \\ f_n(x) = \sum_{k=-3}^{3} \beta_k^n \psi_k^n(x), \end{gathered} \tag{3.19}$$

so that at the higher scales there are the higher frequency oscillations (see Figure 2). It should be also noticed that the lower scale approximations $f_0(x)$, $f_1(x)$, $f_2(x)$ represent the major content of the amplitude. In other words, $f_0(x) + f_1(x) + f_2(x)$ gives a good representation of (3.18) in the origin, while $f_3(x)$, with its higher oscillations, makes a good approximation of the tails of (3.18). Therefore, if we are interested in the evolution of the peak in the origin, we can restrict ourselves to the analysis of the lower scales. If we are interested in the evolution either of the tails or the high frequency, we must take into consideration the higher scales (in our case $f_3(x)$).

If we compare the Shannon wavelet reconstruction with the Fourier integral approach, in the Fourier method the following hold.

(1) It is impossible to have a series expansion except for the periodic functions.

(2) It is impossible to focus, as it is done with the Shannon series, on the contribution of each basis to the function. In other words, the projection of $f(x)$ on each term $\{\cos \xi x, \sin \xi x\}$ of the Fourier basis is not evident. There follows that it is impossible to decompose the profile with the components at different scales.

(3) The integral transform performs an integral over the whole real axis for a function which is substantially zero (over $\mathbb{R}$), except in the "small" interval $(-\varepsilon, \varepsilon)$.

As a second example, let us consider the approximate wavelet representation of the odd function

$$f(x) = e^{-(16x)^2/2} + e^{-4x^2} \sin 2\pi x. \tag{3.20}$$

The bottom length (i.e., the main part) of the function $f(x)$ is concentrated in the interval $[-0.2, 0.2]$. Also in this case, for a localized function, with a low scale $n = 3$ we can have a good approximation (Figures 3, 4) of the function even with a small number k of translation. However, in this case, the error can be reduced (around the origin) by adding some translated instances, but it remains nearly constant far from the origin. In fact, with

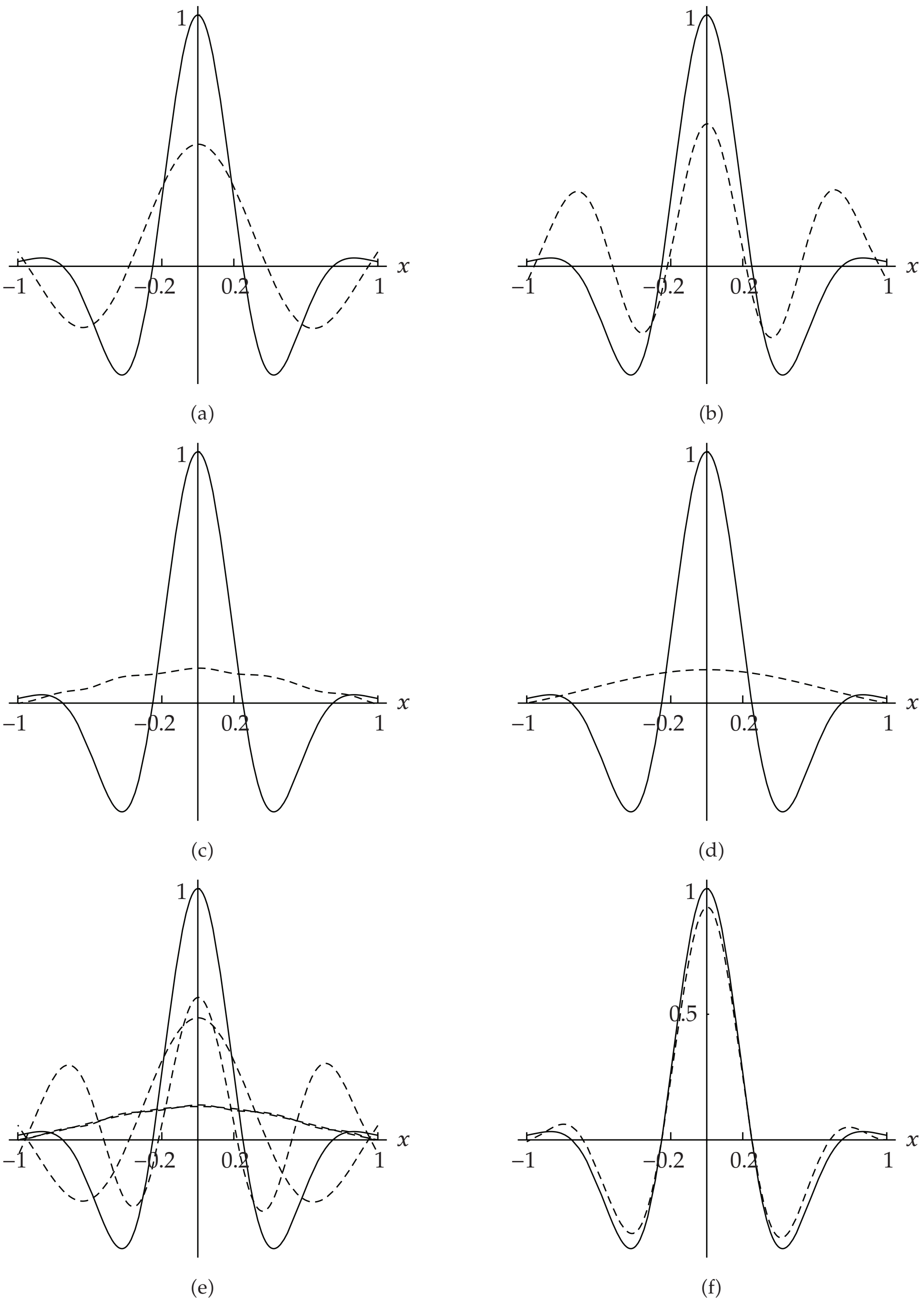


Figure 2: Shannon wavelet reconstruction (dashed) of the even function $f(x) = e^{-4x^2} \cos 2\pi x$, with $n_{\max} = 3$, $-3 \le k \le 3$ (bottom right). Scale approximation with (a) $n = 0$, $-3 \le k \le 3$, (b) $n = 1$, $-3 \le k \le 3$, (c) $n = 2$, $-3 \le k \le 3$, (d) $n = 3$, $-3 \le k \le 3$, (e) $0 \le n \le 3$, $-3 \le k \le 3$, (f) $n = 3$, $-5 \le k \le 5$.

$|k| \le 3$ the absolute value of the approximation error is less than 10% (8% in the origin, Figure 4). The higher number of the translation parameter k improves the approximation of the function on its "tails," in the sense that by increasing the number of translation

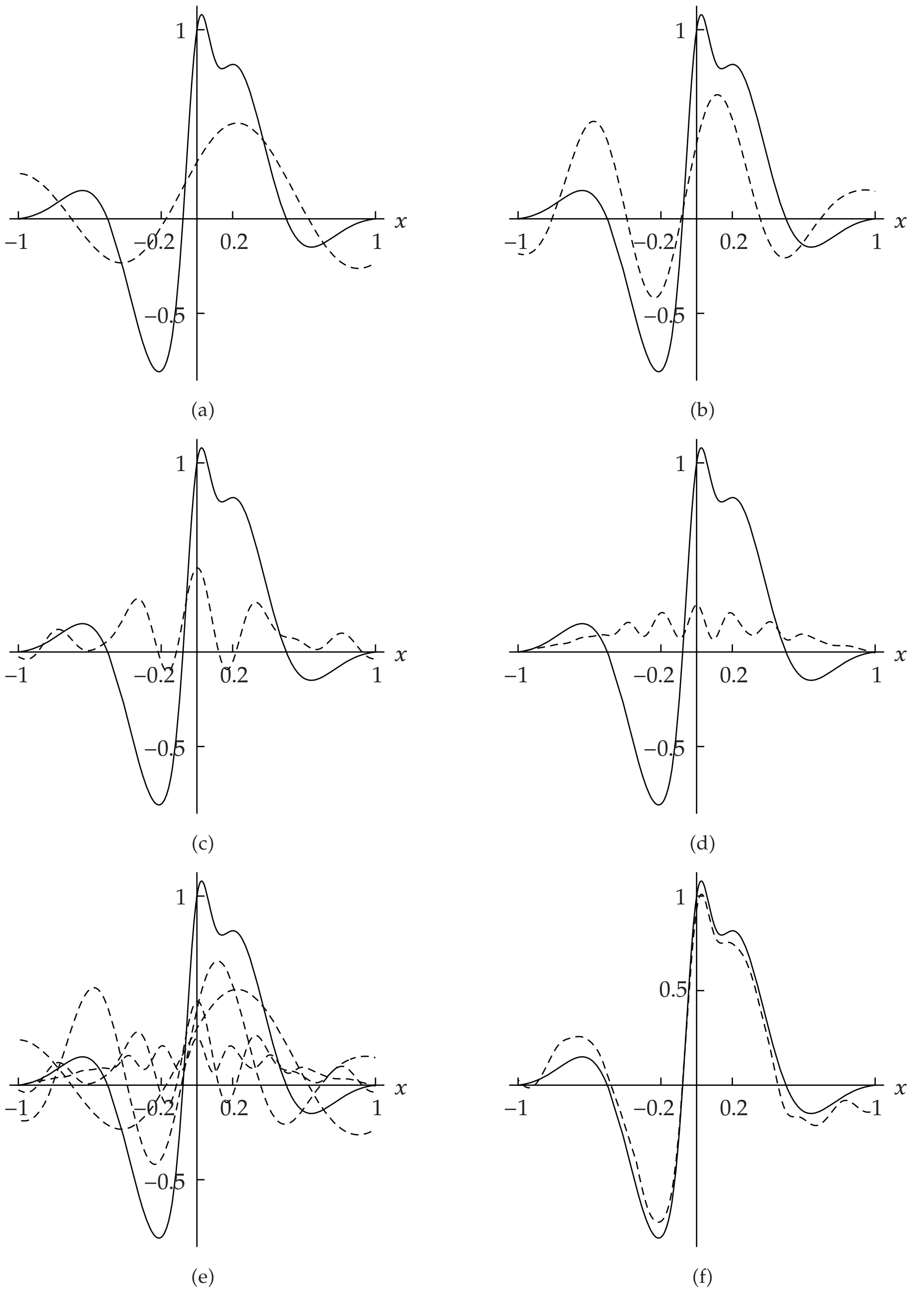
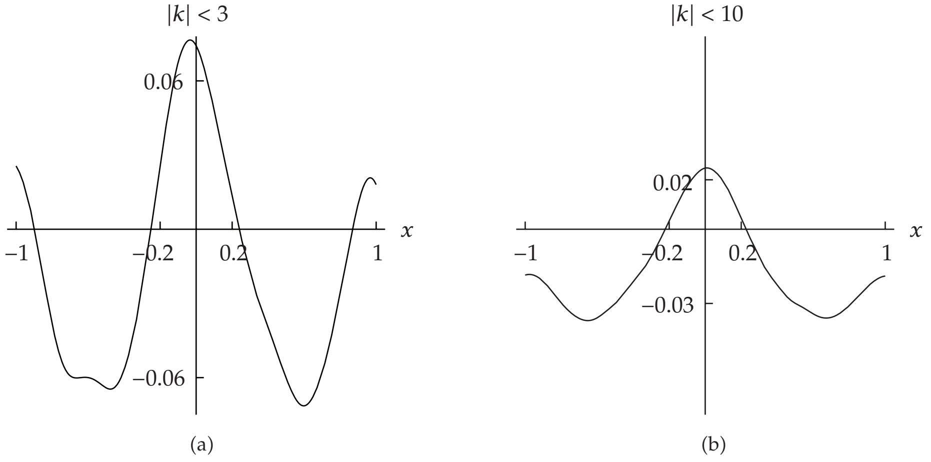


Figure 3: Shannon wavelet reconstruction (dashed) of the odd function $f(x) = e^{-(16x)^2/2} + e^{-4x^2} \sin 2\pi x$, with $N = n_{\max} = 3$, $-3 \le k \le 3$ (bottom right). Scale approximation with (a) $n = 0$, $-3 \le k \le 3$, (b) $n = 1$, $-3 \le k \le 3$, (c) $n = 2$, $-3 \le k \le 3$, (d) $n = 3$, $-3 \le k \le 3$, (e) $0 \le n \le 3$, $-3 \le k \le 3$, (f) $n = 3$, $-5 \le k \le 5$.

parameters$k$ the oscillation on "tails" is reduced and becomes constant (around 10%). But we can see that with $|k| \le 10$ the approximation error in the origin is reduced up to 3%.



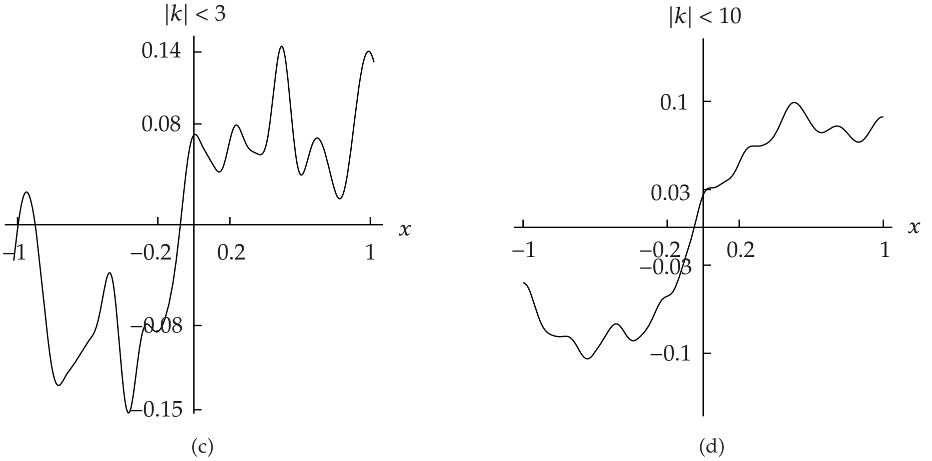


Figure 4: Error of the Shannon wavelet reconstruction of the even function (top) $f(x) = e^{-4x^2}\cos 2\pi x$, with $N = n_{\max} = 3$ and the odd function $f(x) = e^{-(16x)^2/2} + e^{-4x^2}\sin 2\pi x$, with $N = n_{\max} = 3$ (bottom right) with different values of $k_{\max}$.

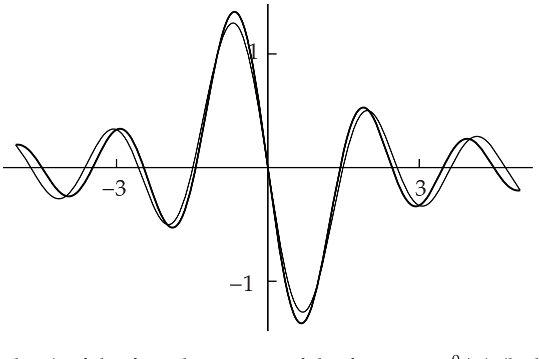


Figure 5: Approximation (plain) of the first derivative of the function $\varphi_0^0(x)$ (bold) by using the connection coefficients.

4. Reconstruction of the Derivatives

Let $f(x) \in L_2(\mathbb{R})$ and let $f(x)$ be a differentiable function $f(x) \in C^p$ with p sufficiently high. The reconstruction of a function $f(x)$ given by (3.10) enables us to compute also its derivatives in terms of the wavelet decomposition

$$\frac{d^\ell}{dx^\ell} f(x) = \sum_{h=-\infty}^{\infty} \alpha_h \frac{d^\ell}{dx^\ell} \varphi_h^0(x) + \sum_{n=0}^{\infty} \sum_{k=-\infty}^{\infty} \beta_k^n \frac{d^\ell}{dx^\ell} \psi_k^n(x), \tag{4.1}$$

so that, according to (3.10), the derivatives of $f(x)$ are known when the derivatives

$$\frac{d^\ell}{dx^\ell} \varphi_h^0(x), \qquad \frac{d^\ell}{dx^\ell} \psi_k^n(x) \tag{4.2}$$

are given.

By a direct computation, we can easily evaluate the first and second derivatives of the scaling function

$$\begin{aligned} \frac{d}{dx} \varphi_k^n(x) &= \frac{2^n\left[-1 + (2^n x - k)\pi \cot((2^n x - k)\pi)\right]}{2^n x - k} \varphi_k^n(x), \\ \frac{d^2}{dx^2} \varphi_k^n(x) &= \frac{2^{2n}\left\{2 - [\pi(2^n x - k)]^2 - 2\pi(2^n x - k)\cot((2^n x - k)\pi)\right\}}{(2^n x - k)^2} \varphi_k^n(x), \end{aligned} \tag{4.3}$$

respectively. However, on this way, higher-order derivatives cannot be easily expressed. Indeed, according to (3.10), we have to compute the wavelet decomposition of the derivatives:

$$\begin{aligned} \frac{d^\ell}{dx^\ell} \varphi_h^0(x) &= \sum_{k=-\infty}^{\infty} \lambda_{hk}^{(\ell)} \varphi_k^0(x) + \sum_{n=0}^{\infty} \sum_{k=-\infty}^{\infty} \Lambda_{hk}^{(\ell)n} \psi_k^n(x), \\ \frac{d^\ell}{dx^\ell} \psi_h^m(x) &= \sum_{k=-\infty}^{\infty} \Gamma_{hk}^{(\ell)m} \varphi_k^0(x) + \sum_{n=0}^{\infty} \sum_{k=-\infty}^{\infty} \gamma_{hk}^{(\ell)mn} \psi_k^n(x), \end{aligned} \tag{4.4}$$

with

$$\lambda_{kh}^{(\ell)} \overset{\text{def}}{\equiv} \left\langle \frac{d^\ell}{dx^\ell} \varphi_k^0(x), \varphi_h^0(x) \right\rangle, \qquad \gamma_{kh}^{(\ell)nm} \overset{\text{def}}{\equiv} \left\langle \frac{d^\ell}{dx^\ell} \psi_k^n(x), \psi_h^m(x) \right\rangle, \tag{4.5}$$

$$\Lambda_{kh}^{(\ell)n} \overset{\text{def}}{\equiv} \left\langle \frac{d^\ell}{dx^\ell} \varphi_k^0(x), \psi_h^n(x) \right\rangle, \qquad \Gamma_{hk}^{(\ell)m} \overset{\text{def}}{\equiv} \left\langle \frac{d^\ell}{dx^\ell} \psi_h^n(x), \varphi_h^0(x) \right\rangle, \tag{4.6}$$

being the connection coefficients [6–9, 11] (or refinable integrals).

Their computation can be easily performed in the Fourier domain, thanks to equality (2.21). In fact, in the Fourier domain the ℓ-order derivatives of the (scaling) wavelet functions are

$$\widehat{\frac{d^\ell}{dx^\ell} \varphi_k^n(x)} = (i\omega)^\ell \widehat{\varphi}_k^n(\omega), \qquad \widehat{\frac{d^\ell}{dx^\ell} \psi_k^n(x)} = (i\omega)^\ell \widehat{\psi}_k^n(\omega) \tag{4.7}$$

and according to (2.19),

$$\begin{aligned}\widehat{\frac{\mathrm{d}^{\ell}}{\mathrm{d}x^{\ell}}\varphi_k^n(x)} &= (i\omega)^{\ell}\frac{2^{-n/2}}{2\pi}e^{-i\omega k/2^n}\chi\left(\frac{\omega}{2^n}+3\pi\right),\\ \widehat{\frac{\mathrm{d}^{\ell}}{\mathrm{d}x^{\ell}}\psi_k^n(x)} &= -(i\omega)^{\ell}\frac{2^{-n/2}}{2\pi}e^{-i\omega(k+1/2)/2^n}\left[\chi\left(\frac{\omega}{2^{n-1}}\right)+\chi\left(\frac{-\omega}{2^{n-1}}\right)\right].\end{aligned}\tag{4.8}$$

Taking into account (2.21), we can easily compute the connection coefficients in the frequency domain

$$\lambda_{kh}^{(\ell)} = 2\pi\left\langle \widehat{\frac{\mathrm{d}^{\ell}}{\mathrm{d}x^{\ell}}\varphi_k^0(x)}, \widehat{\varphi_h^0(x)}\right\rangle, \qquad \gamma_{kh}^{(\ell)nm} = 2\pi\left\langle \widehat{\frac{\mathrm{d}^{\ell}}{\mathrm{d}x^{\ell}}\psi_k^n(x)}, \widehat{\psi_h^m(x)}\right\rangle, \tag{4.9}$$

with the derivatives given by (4.8).

For the explicit computation, we need some preliminary theorems (for a sketch of the proof see also [11]).

Theorem 4.1. *For given $m \in \mathbb{Z}$, $\ell \in \mathbb{N}$, it is*

$$\int x^{\ell}e^{mx}dx = (1-|\mu(m)|)\frac{x^{\ell+1}}{\ell+1} + \mu(m)\frac{e^{mx}}{|m|^{\ell+1}}\sum_{s=1}^{\ell+1}(-1)^{[1+\mu(m)](2\ell-s+1)/2}\frac{\ell!(|m|x)^{\ell-s+1}}{(\ell-s+1)!} + Const, \tag{4.10}$$

where

$$\mu(m) = \text{ sign } (m) = \begin{cases}1, & m>0,\\ -1, & m<0,\\ 0, & m=0.\end{cases}\tag{4.11}$$

Proof. When $m = 0$, (4.10) trivially follows. When $m \neq 0$, by a partial integration we get the iterative formula

$$\int x^{\ell}e^{mx}\mathrm{d}x = \begin{cases}\mu(m)\dfrac{1}{|m|}e^{mx}, & \ell = 0,\\[2ex] \mu(m)\dfrac{1}{|m|}\left[x^{\ell}e^{mx} - \ell\displaystyle\int x^{\ell-1}e^{mx}\mathrm{d}x\right], & \ell > 0,\end{cases}\tag{4.12}$$

from where by the explicit computation of iterative terms and rearranging the many terms, (4.10) holds. □

The following corollary follows. From Theorem 4.1, after a substitution $x \to i\xi$, we have the following corollary.

Corollary 4.2. *For given $m \in \mathbb{Z}$, $\ell \in \mathbb{N}$, it is*

$$\int (i\xi)^{\ell}e^{im\xi}d\xi = i^{\ell}(1-|\mu(m)|)\frac{\xi^{\ell+1}}{\ell+1} - i\mu(m)e^{im\xi}\sum_{s=1}^{\ell+1}(-1)^{[1+\mu(m)](2\ell-s+1)/2}\frac{\ell!(i\xi)^{\ell-s+1}}{(\ell-s+1)!|m|^{s}} + Const. \tag{4.13}$$

In particular, taking into account that

$$e^{ik\pi} = (-1)^k = \begin{cases} 1, & k = \pm 2s, \\ -1, & k = \pm(2s+1), \quad s \in \mathbb{N}, \end{cases} \tag{4.14}$$

we have the following corollary.

Corollary 4.3. *For given $m \in \mathbb{Z} \cup \{0\}$, $\ell \in \mathbb{N}$, and $n \in \mathbb{N}$, it is*

$$\begin{aligned} \int_{-n\pi}^{n\pi} (i\xi)^{\ell} e^{im\xi} d\xi = {} & i^{\ell}(1 - |\mu(m)|) \frac{(n\pi)^{\ell+1}\left[1 + (-1)^{\ell}\right]}{\ell + 1} \\ & + i\mu(m)(-1)^{mn+1} \sum_{s=1}^{\ell+1} (-1)^{[1+\mu(m)](2\ell-s+1)/2} \frac{\ell!(in\pi)^{\ell-s+1}}{(\ell - s + 1)!|m|^s}\left[1 - (-1)^{\ell-s+1}\right]. \end{aligned} \tag{4.15}$$

More in general, the following corollary holds.

Corollary 4.4. *For given $m \in \mathbb{Z}$, $\ell \in \mathbb{N}$, and $a, b \in \mathbb{Z}$ $(a < b)$, it is*

$$\begin{aligned} \int_{a\pi}^{b\pi} (i\xi)^{\ell} e^{im\xi} d\xi = {} & i^{\ell}(1 - |\mu(m)|) \frac{\pi^{\ell+1}\left(b^{\ell+1} - a^{\ell+1}\right)}{\ell + 1} \\ & - i\mu(m) \sum_{s=1}^{\ell+1} (-1)^{[1+\mu(m)](2\ell-s+1)/2} \frac{\ell!(i\pi)^{\ell-s+1}}{(\ell - s + 1)!|m|^s}\left[(-1)^{mb} b^{\ell-s+1} - (-1)^{ma} a^{\ell-s+1}\right]. \end{aligned} \tag{4.16}$$

As a particular case, the following corollaries hold.

Corollary 4.5. *For given $m \in \mathbb{Z}$, $\ell \in \mathbb{N}$, and $b \in \mathbb{Z}$ $(0 < b)$, it is*

$$\begin{aligned} \int_{0}^{b\pi} (i\xi)^{\ell} e^{im\xi} d\xi = {} & i^{\ell}(1 - |\mu(m)|) \frac{\pi^{\ell+1} b^{\ell+1}}{\ell + 1} \\ & - i\mu(m) \left[\sum_{s=1}^{\ell} (-1)^{[1+\mu(m)](2\ell-s+1)/2} \frac{\ell!(i\pi)^{\ell-s+1}(-1)^{mb} b^{\ell-s+1}}{(\ell - s + 1)!|m|^s} \right. \\ & \left. + \frac{(-1)^{(1+\mu(m))\ell/2} \ell!\left[(-1)^{mb} - 1\right]}{|m|^{\ell+1}} \right]. \end{aligned} \tag{4.17}$$

Corollary 4.6. *For given $m \in \mathbb{Z}$, $\ell \in \mathbb{N}$, it is*

$$\int_{0}^{2\pi} (i\xi)^{\ell} e^{im\xi} d\xi = i^{\ell}(1 - |\mu(m)|) \frac{(2\pi)^{\ell+1}}{\ell + 1} - i\mu(m) \sum_{s=1}^{\ell} (-1)^{[1+\mu(m)](2\ell-s+1)/2} \frac{\ell!(2i\pi)^{\ell-s+1}}{(\ell - s + 1)!|m|^s}. \tag{4.18}$$

Thus we can show that the following theorem holds.

Theorem 4.7. *The any order connection coefficients* $(4.5)_1$*of the scaling functions* $\varphi_k^0(x)$ *are*

$$\lambda_{kh}^{(\ell)} = \begin{cases} (-1)^{k-h}\dfrac{i^{\ell}}{2\pi}\displaystyle\sum_{s=1}^{\ell}\dfrac{\ell!\pi^{s}}{s![i(k-h)]^{\ell-s+1}}\left[(-1)^{s}-1\right], & k \neq h, \\ \dfrac{i^{\ell}\pi^{\ell+1}}{2\pi(\ell+1)}\left[1+(-1)^{\ell}\right], & k = h\,, \end{cases} \tag{4.19}$$

or, shortly,

$$\lambda_{kh}^{(\ell)} = \frac{i^{\ell}\pi^{\ell}}{2(\ell+1)}\left[1+(-1)^{\ell}\right](1-|\mu(k-h)|)+(-1)^{k-h}|\mu(k-h)|\frac{i^{\ell}}{2\pi}\sum_{s=1}^{\ell}\frac{\ell!\pi^{s}}{s![i(k-h)]^{\ell-s+1}}\left[(-1)^{s}-1\right]. \tag{4.20}$$

Proof. From (4.9), (4.8), (4.7), (2.21), (2.19), it is

$$\lambda_{kh}^{(\ell)} = \frac{1}{2\pi}\int_{-\infty}^{\infty}(i\omega)^{\ell}e^{-i(k-h)\omega}\chi(\omega+3\pi)\chi(\omega+3\pi)\mathrm{d}\omega, \tag{4.21}$$

that is,

$$\begin{aligned} \lambda_{kh}^{(\ell)} &= \frac{1}{2\pi}\int_{-\infty}^{\infty}(i\omega)^{\ell}e^{-i(k-h)\omega}\chi(\omega+3\pi)\chi(\omega+3\pi)\mathrm{d}\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}(i\omega)^{\ell}e^{-i(k-h)\omega}\mathrm{d}\omega = \frac{i^{\ell}}{2\pi}\int_{-\pi}^{\pi}\omega^{\ell}e^{-i(k-h)\omega}\mathrm{d}\omega. \end{aligned} \tag{4.22}$$

The last integral, according to (4.15) (with $n = 1$), gives (4.20). □

Thus we have at the lower-order derivatives $\ell \leq 5$

$$\begin{aligned} &\lambda_{kh}^{(1)} = -\frac{(-1)^{k-h}}{k-h}, && \lambda_{00}^{(1)} = 0, \\ &\lambda_{kh}^{(2)} = -\frac{2(-1)^{k-h}}{(k-h)^{2}}, && \lambda_{00}^{(2)} = -\frac{\pi^{2}}{3}, \\ &\lambda_{kh}^{(3)} = (-1)^{k-h}\frac{(k-h)^{2}\pi^{2}-6}{(k-h)^{3}}, && \lambda_{00}^{(3)} = 0, \\ &\lambda_{kh}^{(4)} = 4(-1)^{k-h}\frac{(k-h)^{2}\pi^{2}-6}{(k-h)^{4}}, && \lambda_{00}^{(4)} = \frac{\pi^{4}}{5}, \\ &\lambda_{kh}^{(5)} = (-1)^{k-h}\frac{(k-h)^{4}\pi^{4}-20(k-h)^{2}\pi^{2}+120}{(k-h)^{5}}, && \lambda_{00}^{(5)} = 0\,. \end{aligned} \tag{4.23}$$

Analogously for the connection coefficients $(4.5)_2$, we have the following theorem.

Theorem 4.8. *The any order connection coefficients* $(4.5)_2$ *of the Shannon wavelets* $(2.18)_2$ *are*

$$\begin{aligned}\gamma_{kh}^{(\ell)nm} = \delta^{nm}\Bigg\{ & i^{\ell}(1-|\mu(h-k)|)\frac{\pi^{\ell}2^{n\ell-1}}{\ell+1}\left(2^{\ell+1}-1\right)\left(1+(-1)^{\ell}\right) \\ & +\mu(h-k)\sum_{s=1}^{\ell+1}(-1)^{[1+\mu(h-k)](2\ell-s+1)/2}\frac{\ell!\, i^{\ell-s}\pi^{\ell-s}}{(\ell-s+1)!|h-k|^{s}}(-1)^{-s-2(h+k)}2^{n\ell-s-1} \\ & \times\left\{2^{\ell+1}\left[(-1)^{4h+s}+(-1)^{4k+\ell}\right]-2^{s}\left[(-1)^{3k+h+\ell}+(-1)^{3h+k+s}\right]\right\}\Bigg\},\end{aligned} \tag{4.24}$$

respectively, for $\ell \geq 1$, *and* $\gamma_{kh}^{(0)nm} = \delta_{kh}\delta^{nm}$.

Proof. From (4.9), (4.8), (4.7), (2.21), (2.19), it is

$$\begin{aligned}\gamma_{kh}^{(\ell)nm} & \overset{\text{def}}{=} \left\langle \frac{\mathrm{d}^{\ell}}{\mathrm{d}x^{\ell}}\psi_k^n(x), \psi_h^m(x)\right\rangle \\ & \overset{(4.9)}{=} 2\pi\left\langle \widehat{\frac{\mathrm{d}^{\ell}}{\mathrm{d}x^{\ell}}\psi_k^n(x)}, \widehat{\psi_h^m(x)}\right\rangle \\ & \overset{(4.7)}{=} 2\pi\left\langle (i\omega)^{\ell}\widehat{\psi}_k^n(\omega)\,, \widehat{\psi}_h^m(\omega)\right\rangle \\ & \overset{(2.21)}{=} 2\pi\int_{-\infty}^{\infty}(i\omega)^{\ell}\psi_k^n(\omega)\overline{\widehat{\psi}_h^m(\omega)}\mathrm{d}\omega \\ & \overset{(2.19)}{=} 2\pi\int_{-\infty}^{\infty}(i\omega)^{\ell}\frac{2^{-n/2}}{2\pi}e^{-i\omega(k+1/2)/2^n}\left[\chi\left(\frac{\omega}{2^{n-1}}\right)+\chi\left(\frac{-\omega}{2^{n-1}}\right)\right] \\ & \qquad\qquad \times\frac{2^{-m/2}}{2\pi}e^{i\omega(h+1/2)/2^m}\left[\chi\left(\frac{\omega}{2^{m-1}}\right)+\chi\left(\frac{-\omega}{2^{m-1}}\right)\right]\mathrm{d}\omega,\end{aligned} \tag{4.25}$$

from where, according to the definition (2.7), it is

$$\gamma_{kh}^{(\ell)nm} = 0, \quad n \neq m, \tag{4.26}$$

and (for $n = m$)

$$\begin{aligned}\gamma_{kh}^{(\ell)nn} & = \frac{2^{-n}}{2\pi}\int_{-\infty}^{\infty}(i\omega)^{\ell}e^{-i\omega(k-h)/2^n}\left[\chi\left(\frac{\omega}{2^{n-1}}\right)+\chi\left(\frac{-\omega}{2^{n-1}}\right)\right]\mathrm{d}\omega \\ & = \frac{2^{-n}}{2\pi}\left[\int_{-2^{n+1}\pi}^{-2^n\pi}(i\omega)^{\ell}e^{-i\omega(k-h)/2^n}\mathrm{d}\omega + \int_{2^n\pi}^{2^{n+1}\pi}(i\omega)^{\ell}e^{-i\omega(k-h)/2^n}\mathrm{d}\omega\right].\end{aligned} \tag{4.27}$$

By taking into account (4.16), (4.24) is proven. □

Theorem 4.9. *The connection coefficients are recursively given by the matrix at the lowest scale level:*

$$\gamma_{kh}^{(\ell)nn} = 2^{\ell(n-1)}\gamma_{kh}^{(\ell)11}. \tag{4.28}$$

Moreover, it is

$$\gamma_{kh}^{(2\ell+1)nn} = -\gamma_{hk}^{(2\ell+1)nn}, \qquad \gamma_{kh}^{(2\ell)nn} = \gamma_{hk}^{(2\ell)nn}. \tag{4.29}$$

Let us now prove that the mixed connection coefficients (4.6) are zero. It is enough to show the following theorem.

Theorem 4.10. *The mixed coefficients* $(4.6)_1$ *of the Shannon wavelets are*

$$\Lambda_{kh}^{(\ell)n} = 0. \tag{4.30}$$

Proof. From (4.9), (4.8), (4.7), (2.21), (2.19), it is

$$\begin{aligned}
\Lambda_{kh}^{(\ell)n} &\overset{\text{def}}{=} \left\langle \frac{\mathrm{d}^\ell}{\mathrm{d}x^\ell}\varphi_k^0(x), \psi_h^m(x) \right\rangle = 2\pi \left\langle \widehat{\frac{\mathrm{d}^\ell}{\mathrm{d}x^\ell}\varphi_k^0(x)}, \widehat{\psi_h^m(x)} \right\rangle \\
&\overset{(4.7)}{=} 2\pi \left\langle (i\omega)^\ell \widehat{\varphi}_k^0(\omega), \widehat{\psi}_h^m(\omega) \right\rangle \overset{(2.21)}{=} 2\pi \int_{-\infty}^{\infty} (i\omega)^\ell \varphi_k^0(\omega) \overline{\widehat{\psi}_h^m(\omega)} \mathrm{d}\omega \\
&\overset{(2.19)}{=} 2\pi \int_{-\infty}^{\infty} (i\omega)^\ell \frac{2^{-n/2}}{2\pi} e^{-i\omega k/2^n} \chi\left(\frac{\omega}{2^n} + 3\pi\right) \\
&\qquad \times \frac{2^{-m/2}}{2\pi} e^{i\omega(h+1/2)/2^m} \left[\chi\left(\frac{\omega}{2^{m-1}}\right) + \chi\left(\frac{-\omega}{2^{m-1}}\right) \right] \mathrm{d}\omega,
\end{aligned} \tag{4.31}$$

from where, since

$$\chi\left(\frac{\omega}{2^n} + 3\pi\right)\left[\chi\left(\frac{\omega}{2^{m-1}}\right) + \chi\left(\frac{-\omega}{2^{m-1}}\right)\right] = 0, \tag{4.32}$$

the theorem is proven. □

As a consequence, we have that the ℓ-order derivatives of the Shannon scaling and wavelets are

$$\begin{aligned}
\frac{\mathrm{d}^\ell}{\mathrm{d}x^\ell}\varphi_h^0(x) &= \sum_{k=-\infty}^{\infty} \lambda_{hk}^{(\ell)} \varphi_k^0(x), \\
\frac{\mathrm{d}^\ell}{\mathrm{d}x^\ell}\psi_h^m(x) &= \sum_{n=0}^{\infty} \sum_{k=-\infty}^{\infty} \gamma_{hk}^{(\ell)mn} \psi_k^n(x).
\end{aligned} \tag{4.33}$$

In other words, the following theorem holds.

Theorem 4.11. *The derivatives of the Shannon scaling function are orthogonal to the derivatives of the Shannon wavelets*

$$\left\langle \frac{d^\ell}{dx^\ell}\varphi_h^0(x), \frac{d^p}{dx^p}\psi_h^m(x) \right\rangle = 0. \tag{4.34}$$

Proof. It follows directly from (4.33) and the orthogonality of the Shannon functions according to Theorem 2.4. □

4.1. First- and Second-Order Connection Coefficients

For the first and second derivatives of the Shannon wavelets, we have (see [11])

$$\begin{aligned} \frac{d}{dx}\psi_k^n(x) &= \sum_{h=-\infty}^{\infty} \gamma_{kh}^{\prime nn}\psi_h^n(x), \\ \frac{d^2}{dx^2}\psi_k^n(x) &= \sum_{h=-\infty}^{\infty} \gamma_{kh}^{\prime\prime nn}\psi_h^n(x), \end{aligned} \tag{4.35}$$

with (4.24)

$$\begin{aligned} \gamma_{kh}^{\prime nn} &= \mu(h-k)\sum_{s=1}^{2}(-1)^{[1+\mu(h-k)](2-s+1)/2}\frac{i^{1-s}\pi^{1-s}}{(2-s)!|h-k|^s}(-1)^{-s-2(h+k)}2^{n-s-1} \\ &\quad\times\left\{4\left[(-1)^{4h+s}+(-1)^{4k+1}\right]-2^s\left[(-1)^{3k+h+1}+(-1)^{3h+k+s}\right]\right\}, \\ \gamma_{kh}^{\prime\prime nn} &= -(1-|\mu(h-k)|)\pi^2 2^{2n} \\ &\quad+\mu(h-k)\sum_{s=1}^{3}(-1)^{[1+\mu(h-k)](5-s)/2}\frac{2i^{2-s}\pi^{2-s}}{(3-s)!|h-k|^s}(-1)^{-s-2(h+k)}2^{2n-s-1} \\ &\quad\times\left\{8\left[(-1)^{4h+s}+(-1)^{4k+2}\right]-2^s\left[(-1)^{3k+h+2}+(-1)^{3h+k+s}\right]\right\}, \end{aligned} \tag{4.36}$$

respectively.

A disadvantage in (4.33) is that derivatives are expressed as infinite sum. However, since the wavelets are mainly localized in a short range interval, a good approximation can be obtained with a very few terms of the series. The main advantage of (4.33) is that the derivatives are expressed in terms of the wavelet basis.

Analogously, we obtain for the first and second derivative of the scaling function

$$\begin{aligned} \frac{d}{dx}\varphi_k^0(x) &= \sum_{h=-\infty}^{\infty} \lambda'_{kh}\varphi_h^0(x), \\ \frac{d^2}{dx^2}\varphi_k^0(x) &= \sum_{h=-\infty}^{\infty} \lambda''_{kh}\varphi_h^0(x), \end{aligned} \tag{4.37}$$

with (4.20)

$$\begin{aligned} \lambda'_{kh} &= (-1)^{h-k}\mu(h-k)\sum_{s=1}^{2}(-1)^{[1+\mu(h-k)](3-s)/2}\frac{i^{1-s}\pi^{1-s}}{2(2-s)!|h-k|^s}\left[1+(-1)^{1-s}\right], \\ \lambda''_{kh} &= -(1-|\mu(h-k)|)\frac{\pi^2}{2}+(-1)^{h-k}\mu(h-k)\sum_{s=1}^{3}(-1)^{[1+\mu(h-k)](5-s)/2}\frac{i^{2-s}\pi^{2-s}}{(3-s)!|h-k|^s}\left[1+(-1)^{2-s}\right]. \end{aligned} \tag{4.38}$$

The coefficients of derivatives are real values as can be shown by a direct computation

γ'^{11}_{kh}	$k=-2$	$k=-1$	$k=0$	$k=1$	$k=2$
$h=-2$	0	$-\frac{1}{2}$	$-\frac{1}{4}$	$-\frac{1}{6}$	$-\frac{1}{8}$
$h=-1$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{1}{4}$	$-\frac{1}{6}$
$h=0$	$\frac{1}{4}$	$\frac{1}{2}$	0	$-\frac{1}{2}$	$-\frac{1}{4}$
$h=1$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
$h=2$	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$	0

(4.39)

γ'^{22}_{kh}	$k=-2$	$k=-1$	$k=0$	$k=1$	$k=2$
$h=-2$	0	-1	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{4}$
$h=-1$	1	0	-1	$-\frac{1}{2}$	$-\frac{1}{3}$
$h=0$	$\frac{1}{2}$	1	0	-1	$-\frac{1}{2}$
$h=1$	$\frac{1}{3}$	$\frac{1}{2}$	1	0	-1
$h=2$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1	0

(4.40)

If we consider a dyadic discretization of the x-axis such that

$$x_k = 2^{-n}\left(k+\frac{1}{2}\right), \quad k \in \mathbb{Z}, \tag{4.41}$$

that is,

	$k=-2$	$k=-1$	$k=0$	$k=1$	$k=2$
$n=0$	-1.5	-0.5	0.5	1.5	2.5
$n=1$	-0.75	-0.25	0.25	0.75	1.25
$n=2$	-0.375	-0.125	0.125	0.375	0.625

(4.42)

there results

$$\varphi_k^n\left(2^{-n}\left(k+\frac{1}{2}\right)\right)=-2^{n/2},\quad k\in\mathbb{Z}. \tag{4.43}$$

Thus (4.33) at dyadic points x_k= $2^{-n}(k+1/2)$becomes

$$\begin{aligned}\left[\frac{d}{dx}\varphi_k^n(x)\right]_{x=x_k}&=-2^{n/2}\sum_{h=-\infty}^{\infty}\gamma_{kh}^{nn},\\ \left[\frac{d^2}{dx^2}\varphi_k^n(x)\right]_{x=x_k}&=-2^{n/2}\sum_{h=-\infty}^{\infty}\Gamma_{kh}^{nn}.\end{aligned} \tag{4.44}$$

For instance, (see the above tables) in x_1= $2^{-1}(1+1/2)$,

$$\left[\frac{d}{dx}\varphi_1^1(x)\right]_{x=x_1=3/4}=-2^{1/2}\sum_{h=-\infty}^{\infty}\gamma_{1h}^{11}\cong-2^{1/2}\sum_{h=-2}^{2}\gamma_{1h}^{11}=-2^{1/2}\left(\frac{1}{6}+\frac{1}{4}\right)=-\frac{5\sqrt{2}}{12}. \tag{4.45}$$

Analogously, it is

$$\varphi_k^n\left(2^{-n}\left(k+\frac{1}{2}\right)\right)=\frac{2^{1+n/2}}{\pi},\quad k\in\mathbb{Z}, \tag{4.46}$$

from where, in $x_k=(k+1/2)$, it is

$$\begin{aligned}\left[\frac{d}{dx}\varphi_k^0(x)\right]_{x=x_k}&=\frac{2}{\pi}\sum_{h=-\infty}^{\infty}\lambda_{kh},\\ \left[\frac{d^2}{dx^2}\varphi_k^0(x)\right]_{x=x_k}&=\frac{2}{\pi}\sum_{h=-\infty}^{\infty}\Lambda_{kh}.\end{aligned} \tag{4.47}$$

Outside the dyadic points, the approximation is quite good even with low values of the parameters n, k. For instance, we have (Figure 5) the approximation

$$\frac{\mathrm{d}}{\mathrm{d}x}\varphi_0^0(x)=\frac{\cos\pi x}{x}-\frac{\sin\pi x}{\pi x^2}\cong\sum_{h=-5}^{5}\lambda_{0h}\varphi_h^0(x). \tag{4.48}$$

5. Conclusion

In this paper, the theory of Shannon wavelets has been analyzed showing the main properties of these functions sharply localized in frequency. The reconstruction formula for the $L_2(\mathbb{R})$ functions has been given not only for the function but also for its derivatives. The derivative of the Shannon wavelets has been computed by a finite formula (both for the scaling and for the wavelet) for any order derivative. Indeed, to achieve this task, it was enough to compute connection coefficients, that is, the wavelet coefficients of the basis derivatives. These coefficients were obtained as a finite series (for any order derivatives). In Latto's method [6, 8, 9], instead, these coefficients were obtained only (for the Daubechies wavelets) by using the inclusion axiom but in approximated form and only for the first two order derivatives. The knowledge of the derivatives of the basis enables us to approximate a function and its derivatives and it is an expedient tool for the projection of differential operators in the numerical computation of the solution of both partial and ordinary differential equations [2, 3, 10, 13].

References

[1] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, Pa, USA, 1992.

[2] C. Cattani, "Harmonic wavelets towards the solution of nonlinear PDE," *Computers & Mathematics with Applications*, vol. 50, no. 8-9, pp. 1191–1210, 2005.

[3] C. Cattani and J. Rushchitsky, *Wavelet and Wave Analysis as Applied to Materials with Micro or Nanostructure*, vol. 74 of *Series on Advances in Mathematics for Applied Sciences*, World Scientific, Hackensack, NJ, USA, 2007.

[4] G. Toma, "Practical test-functions generated by computer algorithms," in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA '05)*, vol. 3482 of *Lecture Notes in Computer Science*, pp. 576–584, Singapore, May 2005.

[5] M. Unser, "Sampling—50 years after Shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, 2000.

[6] A. Latto, H. L. Resnikoff, and E. Tenenbaum, "The evaluation of connection coefficients of compactly supported wavelets," in *Proceedings of the French-USA Workshop on Wavelets and Turbulence*, Y. Maday, Ed., pp. 76–89, Springer, Princeton, NY, USA, June 1991.

[7] E. B. Lin and X. Zhou, "Connection coefficients on an interval and wavelet solutions of Burgers equation," *Journal of Computational and Applied Mathematics*, vol. 135, no. 1, pp. 63–78, 2001.

[8] J. M. Restrepo and G. K. Leaf, "Wavelet-Galerkin discretization of hyperbolic equations," *Journal of Computational Physics*, vol. 122, no. 1, pp. 118–128, 1995.

[9] C. H. Romine and B. W. Peyton, "Computing connection coefficients of compactly supported wavelets on bounded intervals," Tech. Rep. ORNL/TM-13413, Computer Science and Mathematical Division, Mathematical Sciences Section, Oak Ridge National Laboratory, Oak Ridge, Tenn, USA, http://www.ornl.gov/~webworks/cpr/rpt/91836.pdf.

[10] C. Cattani, "Harmonic wavelet solutions of the Schr"odinger equation," *International Journal of Fluid Mechanics Research*, vol. 30, no. 5, pp. 1–10, 2003.

[11] C. Cattani, "Connection coefficients of Shannon wavelets," *Mathematical Modelling and Analysis*, vol. 11, no. 2, pp. 117–132, 2006.

[12] C. Cattani, "Shannon wavelet analysis," in *Proceedings of the 7th International Conference on Computational Science (ICCS '07)*, Y. Shi, G. D. van Albada, J. Dongarra, and P. M. A. Sloot, Eds., vol. 4488 of *Lecture Notes in Computer Science*, pp. 982–989, Springer, Beijing, China, May 2007.

[13] S. V. Muniandy and I. M. Moroz, "Galerkin modelling of the Burgers equation using harmonic wavelets," *Physics Letters A*, vol. 235, no. 4, pp. 352–356, 1997.

[14] D. E. Newland, "Harmonic wavelet analysis," *Proceedings of the Royal Society of London A*, vol. 443, no. 1917, pp. 203–225, 1993.

[15] W. H"ardle, G. Kerkyacharian, D. Picard, and A. Tsybakov, *Wavelets, Approximation, and Statistical Applications*, vol. 129 of *Lecture Notes in Statistics*, Springer, New York, NY, USA, 1998.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 323080, 16 pages
doi:10.1155/2008/323080

Research Article

Combined Preorder and Postorder Traversal Algorithm for the Analysis of Singular Systems by Haar Wavelets

Beom-Soo Kim,[1] Il-Joo Shim,[2] Myo-Taeg Lim,[3] and Young-Joong Kim[3]

[1] *School of Mechanical and Aerospace Engineering, Gyeongsang National University, 445 Inpyeong-Dong, Tongyeong, Gyeongnam 650-160, South Korea*
[2] *Department of Automatic System Engineering, Daelim College, 526-7 Bisan-Dong, Anyang, Gyeonggi 431-715, South Korea*
[3] *School of Electrical Engineering, Korea University, 1-5 Anam-dong, Sungbuk-gu, Seoul, 136-701, South Korea*

Correspondence should be addressed to Myo-Taeg Lim, mlim@korea.ac.kr

Received 31 May 2008; Accepted 25 August 2008

Recommended by Carlo Cattani

An efficient computational method is presented for state space analysis of singular systems via Haar wavelets. Singular systems are those in which dynamics are governed by a combination of algebraic and differential equations. The corresponding differential-algebraic matrix equation is converted to a generalized Sylvester matrix equation by using Haar wavelet basis. First, an explicit expression for the inverse of the Haar matrix is presented. Then, using it, we propose a combined preorder and postorder traversal algorithm to solve the generalized Sylvester matrix equation. Finally, the efficiency of the proposed method is discussed by a numerical example.

1. Introduction

Wavelets are mathematical functions that cut up data into different frequency components and then study each component with a resolution matched to its scale. Wavelets are now being applied in many areas of science and engineering [1–4]. Much attention has been focused on the use of wavelet transforms to investigate dynamic systems. This is due to the powerful ability of wavelet transforms to decompose time series in time-frequency domain and wavelet basis functions. Chen and Hsiao [3, 4] derived a Haar operational matrix for integration and solved the lumped and distributed parameter systems by constructing operational matrices of various order. The main characteristic of this technique is that it converts a differential equation into an algebraic one with the result that the solution procedures are greatly reduced and simplified. This approach gives new insight into the use of the Haar wavelet method.

Singular systems (also referred to as descriptor or semistate systems) arise more naturally than do state-variable descriptions in the analysis of many sorts of systems. Examples occur in electrical networks, neural networks, control systems, chemical systems, economic systems, and so on (see [5, 6] and references therein). These systems are governed by a mixture of differential and algebraic equations. The complex nature of singular systems causes many difficulties in the analytical and numerical treatment of such systems.

Recently, Haar wavelet technique was applied to state analysis and observer design of singular systems [7]. This approach replaces the state function and the forcing function by the truncated Haar series, respectively. Then the state trajectories are obtained by solving a generalized Sylvester matrix equation. But there exists a trade-off between the resolution of the wavelets and the computation time. The accuracy of the solution can be achieved by increasing the resolution level, but this requires more computation time and very large memory.

In this paper, an efficient computational method is presented for state space analysis of singular systems via Haar wavelets. First, an explicit expression for the inverse of the Haar matrix is presented. This inverse matrix also has a recursive structure. By using this matrix, we propose a combined preorder and postorder traversal algorithm. Then, the full-order generalized Sylvester matrix equation should be solved in terms of the solutions of simple linear matrix equations. Finally, the efficiency of the proposed method is discussed by a numerical example.

2. Kronecker product

Let $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ be $n \times p$ and $r \times q$ matrices, respectively. The Kronecker product of the matrices, denoted by $\mathbf{A} \otimes \mathbf{B} (\in \mathbb{R}^{nr \times pq})$, is defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1p}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & & a_{2p}\mathbf{B} \\ \vdots & & & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \cdots & a_{np}\mathbf{B} \end{bmatrix}. \tag{2.1}$$

The *vec* operator transforms a matrix $\mathbf{A}$ of size $n \times p$ to a vector of size $np \times 1$ by stacking the columns of $\mathbf{A}$. Some properties of the Kronecker product are given below [8]:

$$\begin{gathered} (\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C}, \\ (\mathbf{A} \otimes \mathbf{B})\mathbf{C} = (\mathbf{AC} \otimes \mathbf{B}), \\ (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD}), \\ (\mathbf{A} \otimes \mathbf{B})^T = \mathbf{A}^T \otimes \mathbf{B}^T. \end{gathered} \tag{2.2}$$

3. Haar wavelets and their properties

Wavelets constitute a family of functions constructed from dilation and translation of a single function called the mother wavelet that generates orthogonal bases of $L_2(R)$. The simplest and most basic of the wavelet systems is the Haar wavelet which is a group of square waves with magnitudes of ±1 in certain intervals and zeros elsewhere [9]. The scaling function $\varphi_0(t)$ and mother wavelet $\varphi_1(t)$ are defined by, respectively,

$$\varphi_0(t) = \begin{cases} 1, & t \in [0,1), \\ 0, & t \notin [0,1), \end{cases}$$
$$\varphi_1(t) = \begin{cases} 1, & t \in \left[0, \dfrac{1}{2}\right), \\ -1, & t \in \left[\dfrac{1}{2}, 1\right), \\ 0, & t \notin [0,1). \end{cases} \tag{3.1}$$

Then, all the other basis functions $\varphi_k(t)$ are obtained by dilation and translation of the mother wavelet as follows:

$$\varphi_k(t) = \varphi_1(2^n t - j) = \begin{cases} 1, & t \in [t_a, t_b), \\ -1, & t \in [t_b, t_c), \\ 0, & t \notin [t_a, t_c), \end{cases} \tag{3.2}$$

where $k = 2^n + j$, integer $n \geq 1$ is a dilation parameter, integer $0 \leq j < 2^n$ is a shift parameter, and the intervals are given by $t_a = m/2^n$, $t_b = (0.5 + j)/2^n$, and $t_c = (1 + j)/2^n$. Since the support of the Haar wavelet is $[0,1)$, any square integrable function $y(t) \in L_2[0,1)$ can be written as an infinite linear combination of Haar functions

$$y(t) = \sum_{k=0}^{\infty} c_k \varphi_k(t), \quad t \in [0,1), \tag{3.3}$$

where the Haar coefficients are determined by

$$c_k = \langle y(t), \varphi_k(t) \rangle = 2^n \int_0^1 y(t) \varphi_k(t) dt, \tag{3.4}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. In practical applications, Haar series are truncated to m terms, that is,

$$y(t) \cong \sum_{k=0}^{m-1} c_k \varphi_k(t) = \mathbf{C}_m^{\mathrm{T}} \mathbf{h}_m(t), \tag{3.5}$$

where Haar functions coefficient vector $\mathbf{C}_m$ and Haar functions vector $\mathbf{h}_m$ are defined as $\mathbf{C}_m \triangleq [c_0 \ \ c_1 \ \ \cdots \ \ c_{m-1}]^{\mathrm{T}}$ and $\mathbf{h}_m(t) \triangleq [\varphi_0(t) \ \ \varphi_1(t) \ \ \cdots \ \ \varphi_{m-1}(t)]^{\mathrm{T}}$.

Integrals of the Haar functions with respect to variable t form ramp and triangular waveforms standing with uniform slope, respectively, at the positions of the corresponding rectangular functions. The group of these integrals can be expressed as follows:

$$\begin{gathered}\int_0^1 \varphi_0(t)dt = t, \quad t \in \left[t_a, t_b\right), \\ \int_0^1 \varphi_k(t)dt = \begin{cases} t - t_a, & t \in \left[t_a, t_b\right), \\ -t + t_c, & t \in \left[t_b, t_c\right), \\ 0, & t \notin \left[t_a, t_c\right). \end{cases}\end{gathered} \tag{3.6}$$

Then, the Haar matrix $\mathbf{H}_m$ is defined as

$$\mathbf{H}_m(t) \triangleq \left[\mathbf{h}_m(t_0) \;\; \mathbf{h}_m(t_1) \;\; \cdots \;\; \mathbf{h}_m(t_{m-1})\right], \tag{3.7}$$

where $i/m \leq t_i \leq (i+1)/m$.

Integration of the Haar function vector can be written as

$$\int_0^t \mathbf{h}_m(\tau)d\tau \cong \mathbf{P}_m \mathbf{h}_m(t), \tag{3.8}$$

where $\mathbf{P}_m$ is the m-square operational matrix of integration which satisfies the following recursive formula [3]:

$$\mathbf{P}_m = \begin{bmatrix} \mathbf{P}_{m/2} & -\dfrac{1}{2m}\mathbf{H}_{m/2} \\ \dfrac{1}{2m}\mathbf{H}_{m/2}^{-1} & \mathbf{0}_{m/2} \end{bmatrix}, \qquad \mathbf{P}_1 = \left[\frac{1}{2}\right], \tag{3.9}$$

where $\mathbf{0}_{m/2}$ is an $m/2$-square zero matrix. The Haar matrix $\mathbf{H}_m$ also has the following recursive formula [3]:

$$\mathbf{H}_m = \begin{bmatrix} \mathbf{H}_{m/2} \otimes \left[1 \;\; 1\right] \\ \mathbf{I}_{m/2} \otimes \left[1 \;\; -1\right] \end{bmatrix}, \qquad \mathbf{H}_1 = [1]. \tag{3.10}$$

Particularly, it was proven that the following relationship holds [3]:

$$\mathbf{H}_m^{-1} = \frac{1}{m}\mathbf{H}_m^T \mathbf{D}_m, \tag{3.11}$$

where $\mathbf{D}_m = \mathrm{diag}(1\ 1\ 2\ 2 \cdots \underbrace{2^{p-1} \cdots 2^{p-1}}_{m/2})$ and $p = \log_2 m$. This diagonal matrix $\mathbf{D}_m$ also can be represented in the recursive form

$$\mathbf{D}_m = \begin{bmatrix} \mathbf{D}_{m/2} & \mathbf{0}_{m/2} \\ \mathbf{0}_{m/2} & \dfrac{m}{2}\mathbf{I}_{m/2} \end{bmatrix}, \qquad \mathbf{D}_1 = [1], \tag{3.12}$$

where $m = 2^k,\ k = 1, 2, \ldots, J$, and J is called a resolution scale or level.

We present the following lemma which will be used to decompose the generalized Sylvester matrix equation.

Lemma 3.1. *Let $\mathbf{H}_m$ be a Haar matrix defined in* (3.10)*. Then, its inverse matrix has the following recursive form:*

$$\mathbf{H}_m^{-1} = \left[\mathbf{H}_{m/2}^{-1} \otimes \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \quad \mathbf{I}_{m/2} \otimes \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}\right]. \tag{3.13}$$

Proof. We assume that $\mathbf{H}_m^{-1}$ has the following recursive structure:

$$\mathbf{H}_m^{-1} = \left[\mathbf{H}_{m/2}^{-1} \otimes \begin{bmatrix} a \\ b \end{bmatrix} \quad \mathbf{I}_{m/2} \otimes \begin{bmatrix} c \\ d \end{bmatrix}\right], \tag{3.14}$$

where $a,\ b,\ c,\ d$ are constants to be determined. Now, we multiply $\mathbf{H}_m$ and $\mathbf{H}_m^{-1}$:

$$\mathbf{H}_m\mathbf{H}_m^{-1} = \begin{bmatrix} \left(\mathbf{H}_{m/2} \otimes [1\ \ 1]\right)\left(\mathbf{H}_{m/2}^{-1} \otimes \begin{bmatrix} a \\ b \end{bmatrix}\right) & \left(\mathbf{H}_{m/2} \otimes [1\ \ 1]\right)\left(\mathbf{I}_{m/2} \otimes \begin{bmatrix} c \\ d \end{bmatrix}\right) \\ \left(\mathbf{I}_{m/2} \otimes [1\ \ -1]\right)\left(\mathbf{H}_{m/2}^{-1} \otimes \begin{bmatrix} a \\ b \end{bmatrix}\right) & \left(\mathbf{I}_{m/2} \otimes [1\ \ -1]\right)\left(\mathbf{I}_{m/2} \otimes \begin{bmatrix} c \\ d \end{bmatrix}\right) \end{bmatrix}. \tag{3.15}$$

Then, using the property of $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, we obtain

$$\mathbf{H}_m\mathbf{H}_m^{-1} = \begin{bmatrix} \mathbf{I}_{m/2} \otimes (a+b) & \mathbf{H}_{m/2} \otimes (c+d) \\ \mathbf{H}_{m/2}^{-1} \otimes (a-b) & \mathbf{I}_{m/2} \otimes (c-d) \end{bmatrix}. \tag{3.16}$$

Thus, $a = b = 0.5,\ c = 0.5,\ d = -0.5$ satisfy $\mathbf{H}_m\mathbf{H}_m^{-1} = \mathbf{H}_m^{-1}\mathbf{H}_m = \mathbf{I}$. □

4. Singular linear system

Consider a linear continuous-time singular system described by

$$\mathbf{E}\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{4.1}$$

where $\mathbf{x}(t) \in \mathbb{R}^p$ denotes the vector of state variables, $\mathbf{u}(t) \in \mathbb{R}^q$ denotes the vector of manipulated inputs, $\mathbf{E}$, $\mathbf{A}$ are $p \times p$ matrices, $\mathbf{E}$ is generally singular, and $\mathbf{B}$ is a $p \times q$

matrix. Without loss of generality, we assume that $\mathrm{rank}(\mathbf{A}) = p$ and (4.1) is regular, that is, $\det(\lambda\mathbf{E} - \mathbf{A}) \neq 0$. Regularity means that the solution $\mathbf{x}(t)$ is uniquely determined by the given initial value $\mathbf{x}_0$ and input $\mathbf{u}(t)$.

If the input function vector $\mathbf{u}(t)$ is square integrable in the interval $[0, 1)$, then it can be represented in a Haar function basis $\mathbf{h}_m(t)$ as

$$\mathbf{u}(t) = \mathbf{G}\mathbf{h}_m(t), \tag{4.2}$$

where $\mathbf{G} \in \mathbb{R}^{q\times m}$ is a Haar coefficient matrix and can be obtained by the method described in Section 3. Likewise, $\dot{\mathbf{x}}(t)$ is expanded in Haar function basis

$$\dot{\mathbf{x}}(t) = \mathbf{V}\mathbf{h}_m(t), \tag{4.3}$$

where $\mathbf{V} \in \mathbb{R}^{p\times m}$ is the unknown matrix to be determined. From the definition of the Haar function, the initial state can be represented as follows:

$$\mathbf{x}_0 = \begin{bmatrix}\mathbf{x}_0 & 0 & \cdots & 0\end{bmatrix}\mathbf{h}_m(t). \tag{4.4}$$

Integrating (4.3) from 0 to t, we have

$$\mathbf{x}(t) = \mathbf{V}\mathbf{P}_m\mathbf{h}_m(t) + \mathbf{x}_0. \tag{4.5}$$

Integrating (4.1) and using (3.8) and (4.4), after canceling $\mathbf{h}_m(t)$, we obtain

$$\mathbf{E}\mathbf{V} - \mathbf{A}\mathbf{V}\mathbf{P}_m = \mathbf{Q}, \tag{4.6}$$

where we define $\mathbf{Q} \triangleq \begin{bmatrix}\mathbf{A}\mathbf{x}_0 & 0 & \cdots & 0\end{bmatrix} + \mathbf{B}\mathbf{G}$. Thus, the differential matrix equation (4.1) has been transformed to a generalized Sylvester matrix equation that must be solved for $\mathbf{V}$. Equation (4.6) can be solved by using Kronecker product as in [6]

$$\left(\mathbf{I}_m \otimes \mathbf{E} + \mathbf{P}_m^{\mathrm{T}} \otimes \mathbf{A}\right)\mathrm{vec}(\mathbf{V}) = \mathrm{vec}(\mathbf{Q}), \tag{4.7}$$

where $\mathbf{I}_m$ is a unit matrix. Equation (4.7) can be solved by LU factorization. However, the coefficient matrix $\mathbf{I}_m \otimes \mathbf{E} + \mathbf{P}_m^{\mathrm{T}} \otimes \mathbf{A}$ has dimension $pm \times pm$, making this approach impractical except for small systems. There are other methods for solving the Sylvester matrix equation (4.6), for example, the Bartels-Stewart algorithm, Krylov subspace method, and matrix sign function method (see [10] and references therein). In [3, 11], recursive algorithms were derived to solve the equations of type $\mathbf{V} - \mathbf{A}\mathbf{V}\mathbf{P}_m = \mathbf{Q}$ for linear systems. It should be noted that the algorithm in [3] is not applicable to a generalized Sylvester matrix equation (4.6), since $\mathbf{E}$ is a singular matrix.

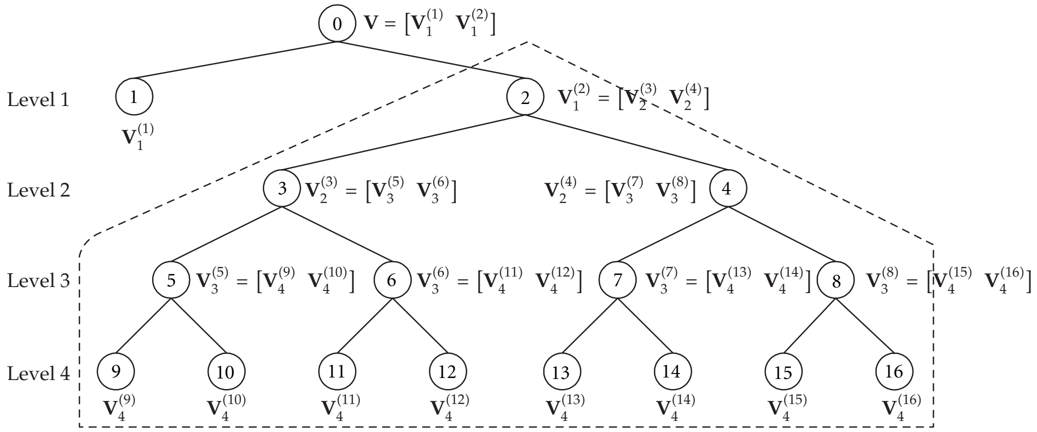

0
V = [V_1^(1) V_1^(2)]
Level 1
1
V_1^(1)
2
V_1^(2) = [V_2^(3) V_2^(4)]
Level 2
3
V_2^(3) = [V_3^(5) V_3^(6)]
V_2^(4) = [V_3^(7) V_3^(8)]
4
Level 3
5
V_3^(5) = [V_4^(9) V_4^(10)]
6
V_3^(6) = [V_4^(11) V_4^(12)]
7
V_3^(7) = [V_4^(13) V_4^(14)]
8
V_3^(8) = [V_4^(15) V_4^(16)]
Level 4
9
10
11
12
13
14
15
16
V_4^(9)
V_4^(10)
V_4^(11)
V_4^(12)
V_4^(13)
V_4^(14)
V_4^(15)
V_4^(16)

Figure 1: Binary tree for resolution scale $J = 4$.

4.1. Decomposition and recursive binary tree

Under the assumption that **A** is a nonsingular matrix, (4.6) can be written as the following Sylvester equation:

$$\mathbf{A}^{-1}\mathbf{E}\mathbf{V} - \mathbf{V}\mathbf{P}_s = \mathbf{A}^{-1}\mathbf{Q}. \tag{4.8}$$

To decompose (4.8), we split $\mathbf{V}$ and $\mathbf{A}^{-1}\mathbf{Q}$ by columns:

$$\mathbf{A}_E\left[\mathbf{V}_1^{(1)} \quad \mathbf{V}_1^{(2)}\right] - \left[\mathbf{V}_1^{(1)} \quad \mathbf{V}_1^{(2)}\right]\begin{bmatrix} \mathbf{P}_{m/2} & -\dfrac{1}{2m}\mathbf{H}_{m/2} \\ \dfrac{1}{2m}\mathbf{H}_{m/2}^{-1} & \mathbf{0}_{m/2} \end{bmatrix} = \left[\mathbf{Q}_a \quad \mathbf{Q}_b\right], \tag{4.9}$$

where $\mathbf{A}_E \triangleq \mathbf{A}^{-1}\mathbf{E}$, $\mathbf{A}^{-1}\mathbf{Q} = \left[\mathbf{Q}_a \quad \mathbf{Q}_b\right]$, $\mathbf{V} = \left[\mathbf{V}_1^{(1)} \quad \mathbf{V}_1^{(2)}\right]$ with $\mathbf{Q}_a, \mathbf{Q}_b, \mathbf{V}_1^{(1)}, \mathbf{V}_1^{(2)} \in R^{p\times m/2}$. Here $\mathbf{V}_k^{(r)}$ denotes the matrix that is decomposed at level k with $r = \{2^k, 2^k - 1\}$. Then, we obtain the following reduced-order matrix equations:

$$\mathbf{A}_E\mathbf{V}_1^{(1)} - \mathbf{V}_1^{(1)}\mathbf{P}_{m/2} - \frac{1}{2m}\mathbf{V}_1^{(2)}\mathbf{H}_{m/2}^{-1} = \mathbf{Q}_a. \tag{4.10}$$

$$\mathbf{A}_E\mathbf{V}_1^{(2)} + \frac{1}{2m}\mathbf{V}_1^{(1)}\mathbf{H}_{m/2} = \mathbf{Q}_b. \tag{4.11}$$

Since **E** is a singular matrix, $\mathbf{A}_E$ is also singular. Thus, we postmultiply by $\mathbf{H}_{m/2}^{-1}$ both sides of (4.11) to express $\mathbf{V}_1^{(1)}$ in terms of $\mathbf{V}_1^{(2)}$

$$\mathbf{V}_1^{(1)} = -2m\mathbf{A}_E\mathbf{V}_1^{(2)}\mathbf{H}_{m/2}^{-1} + 2m\mathbf{Q}_b\mathbf{H}_{m/2}^{-1}. \tag{4.12}$$

Substituting (4.12) into (4.10) yields

$$\begin{aligned}&-2m\mathbf{A}_E^2\mathbf{V}_1^{(2)}\mathbf{H}_{m/2}^{-1}+2m\mathbf{A}_E\mathbf{Q}_b\mathbf{H}_{m/2}^{-1}+2m\mathbf{A}_E\mathbf{V}_1^{(2)}\mathbf{H}_{m/2}^{-1}\mathbf{P}_{m/2}\\&\quad-2m\mathbf{Q}_b\mathbf{H}_{m/2}^{-1}\mathbf{P}_{m/2}-\frac{1}{2m}\mathbf{V}_1^{(2)}\mathbf{H}_{m/2}^{-1}=\mathbf{Q}_a.\end{aligned}\tag{4.13}$$

Therefore, the original problem is decomposed into a reduced-order generalized Sylvester matrix equation (4.13) and a matrix algebraic equation (4.12). Again postmultiplying by $\mathbf{H}_{m/2}$ both sides of (4.13), we have

$$\begin{aligned}&\left(-2m\mathbf{A}_E^2-\frac{1}{2m}\mathbf{I}\right)\mathbf{V}_1^{(2)}+2m\mathbf{A}_E\mathbf{V}_1^{(2)}\mathbf{H}_{m/2}^{-1}\mathbf{P}_{m/2}\mathbf{H}_{m/2}\\&\quad=\mathbf{Q}_a\mathbf{H}_{m/2}-2m\mathbf{A}_E\mathbf{Q}_b+2m\mathbf{Q}_b\mathbf{H}_{m/2}^{-1}\mathbf{P}_{m/2}\mathbf{H}_{m/2}.\end{aligned}\tag{4.14}$$

In (4.14), we define

$$\mathbf{C}_{m/2}\triangleq\mathbf{H}_{m/2}^{-1}\mathbf{P}_{m/2}\mathbf{H}_{m/2}.\tag{4.15}$$

Then, the matrix $\mathbf{C}_{m/2}$ is an upper triangular matrix and has the following recursive form:

$$\mathbf{C}_{m/2}=\begin{bmatrix}\frac{1}{2}\mathbf{C}_{m/4}&\frac{2}{m}\mathbf{1}_{m/4}\\\mathbf{0}_{m/4}&\frac{1}{2}\mathbf{C}_{m/4}\end{bmatrix},\qquad\mathbf{C}_1=\left[\frac{1}{2}\right],\tag{4.16}$$

where $\mathbf{1}_{m/4}$ denotes $m/4$-square matrix with all elements being 1 (see Appendix A).

Substituting (4.16) into (4.14) and splitting $\mathbf{V}_1^{(2)}$ and the right-hand side of (4.14) by columns yields

$$\mathbf{A}_h\left[\mathbf{V}_2^{(3)}\;\;\mathbf{V}_2^{(4)}\right]+2m\mathbf{A}_E\left[\mathbf{V}_2^{(3)}\;\;\mathbf{V}_2^{(4)}\right]\begin{bmatrix}\frac{1}{2}\mathbf{C}_{m/4}&\frac{2}{m}\mathbf{1}_{m/4}\\\mathbf{0}_{m/4}&\frac{1}{2}\mathbf{C}_{m/4}\end{bmatrix}=\left[\mathbf{T}_2^{(3)}\;\;\mathbf{T}_2^{(4)}\right],\tag{4.17}$$

where

$$\begin{aligned}\mathbf{A}_h&\triangleq\left(-2m\mathbf{A}_E^2-\frac{1}{2m}\mathbf{I}\right),\\\mathbf{Q}_a\mathbf{H}_{m/2}-2m\mathbf{A}_E\mathbf{Q}_b+2m\mathbf{Q}_b\mathbf{C}_{m/2}&\triangleq\mathbf{T}_1^{(2)}=\left[\mathbf{T}_2^{(3)}\;\;\mathbf{T}_2^{(4)}\right],\\\mathbf{V}_1^{(2)}&=\left[\mathbf{V}_2^{(3)}\;\;\mathbf{V}_2^{(4)}\right].\end{aligned}\tag{4.18}$$

Thus, (4.17) is decomposed into two matrix equations with dependent and independent subsystems.

$$\mathbf{A}_h\mathbf{V}_2^{(3)} + m\mathbf{A}_E\mathbf{V}_2^{(3)}\mathbf{C}_{m/4} = \mathbf{T}_2^{(3)}. \tag{4.19}$$

$$\mathbf{A}_h\mathbf{V}_2^{(4)} + m\mathbf{A}_E\mathbf{V}_2^{(4)}\mathbf{C}_{m/4} = \mathbf{T}_2^{(4)} - 4\mathbf{A}_E\mathbf{V}_2^{(3)}\mathbf{1}_{m/4}. \tag{4.20}$$

In (4.19) and (4.20), we first solve for $\mathbf{V}_2^{(3)}$ and then after updating the right-hand side of (4.20) with respect to $\mathbf{V}_2^{(3)}$, solve for $\mathbf{V}_2^{(4)}$. Since (4.19) and (4.20) have the same form as (4.17) and $\mathbf{C}_{m/4}$ is still an upper triangular matrix, they can be decomposed into two subsystems in which the dimension has been reduced by half, respectively. Therefore, we recursively decompose each equation into two equations until no further decomposition is possible in which all $\mathbf{V}_J^{(r)}$, $\mathbf{T}_J^{(r)}$ $(r = 2^{J-1} + 1, \ldots, 2^J)$ are column vectors. This procedure constructs the binary tree as shown in Figure 1.

A binary tree is a rooted tree in which each node has at most two children, designated as a left child and a right child. A full binary tree is a binary tree in which each node has exactly two children or none. A perfect (or complete) binary tree is a full binary tree in which all leaves have the same depth [12]. In Figure 1, the binary tree in the dotted box is a perfect binary tree of depth $J - 1$. An external node (or leaf node) is a node with no children. For instance, the nodes labeled 1, 9, 10, 11, 12, 13, 14, 15, and 16 in Figure 1 are external nodes.

Matrix equations corresponding to all external nodes of the perfect binary tree are classified into two types of equations described as follows:

$$\begin{aligned} \mathbf{A}_h\mathbf{V}_J^{(r)} + 4\mathbf{A}_E\mathbf{V}_J^{(r)}\mathbf{C}_1 &= \mathbf{T}_J^{(r)}, \quad r = 2^{J-1}+1,\ 2^{J-1}+3, \ldots, 2^J - 1 \ (r \text{ is odd}), \\ \mathbf{A}_h\mathbf{V}_J^{(r)} + 4\mathbf{A}_E\mathbf{V}_J^{(r)}\mathbf{C}_1 &= \mathbf{T}_J^{(r)} - 4\mathbf{A}_E\mathbf{V}_J^{(r-1)}\mathbf{1}_1, \quad r = 2^{J-1}+2,\ 2^{J-1}+4, \ldots, 2^J \ (r \text{ is even}). \end{aligned} \tag{4.21}$$

Note that in equation (4.21), $\mathbf{C}_1 = 1/2$, $\mathbf{1}_1 = 1$. Thus, they become simple linear matrix equations as follows:

$$\begin{aligned} (\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_J^{(r)} &= \mathbf{T}_J^{(r)}, \quad \text{if } r \text{ is odd}, \\ (\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_J^{(r)} &= \mathbf{T}_J^{(r)} - 4\mathbf{A}_E\mathbf{V}_J^{(r-1)}, \quad \text{if } r \text{ is even}. \end{aligned} \tag{4.22}$$

4.2. Combined preorder and postorder traversal algorithm

Visiting all the nodes in a tree in some particular order is known as a tree traversal. A preorder traversal visits the root of a subtree, then the left and right subtrees recursively. A postorder traversal visits the left and right subtrees recursively, then the root node of the subtree [12]. For example, the preorder and postorder traversals of the binary tree shown in Figure 1 are as follows:

Preorder traversal: 0 → 1 → 2 → 3 → 5 → 9 → 10 → 6 → 7 → 12 → 4 → 7 → 13 → 14 → 8 → 15 → 16

Postorder traversal: 1 → 9 → 10 → 5 → 11 → 12 → 6 → 3 → 13 → 14 → 7 → 15 → 16 → 8 → 4 → 2 → 0

```
Step 1. Initialize $\mathbf{A}_h$, $\mathbf{A}_E$, $\mathbf{T}$.
Step 2. Obtain $\mathbf{V}_1^{(2)}$
Input: Resolution scale $J$
$WaveSolver(J)$
{
  for $(r = 2^{J-1} + 1;\ r < 2^J;\ r = r + 2)$
  {
    Solve for $\mathbf{V}_J^{(r)}$ the system $(\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_J^{(r)} = \mathbf{T}_J^{(r)}$
    Update $\mathbf{T}_J^{(r+1)}$ accroding to $\mathbf{T}_J^{(r+1)} = \mathbf{T}_J^{(r+1)} - 4\mathbf{A}_E\mathbf{V}_J^{(r)}$
    Solve for $\mathbf{V}_J^{(r+1)}$ the system $(\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_J^{(r+1)} = \mathbf{T}_J^{(r+1)}$
    $WaveTree\ (1, J, r + 1)$
  }
}
Input: rno is a number of recursive call.
    Resolution scale $J$
    $r$ is a node number.
$WaveTree(\mathrm{rno}, J, r)$
{
  if $(J - \mathrm{rno} \le 0)$
     return
  Merge: $\mathbf{V}_{J-\mathrm{rno}}^{(r/2)} = \left[\mathbf{V}_{J-\mathrm{rno}+1}^{(r-1)} \quad \mathbf{V}_{J-\mathrm{rno}+1}^{(r)}\right]$
  if $\left(\dfrac{r}{2} \text{ is even}\right)$
    $WaveTree\left(\mathrm{rno}{+}1, J, \dfrac{r}{2}\right)$;
  else
    Update and Split: $\left[\mathbf{T}_{J-\mathrm{rno}+1}^{(r-1)} \quad \mathbf{T}_{J-\mathrm{rno}+1}^{(r)}\right] = \mathbf{T}_{J-\mathrm{rno}}^{(r/2+1)} - 4\mathbf{A}_E\mathbf{V}_{J-\mathrm{rno}}^{(r/2)}\mathbf{1}_{m/2^{J-\mathrm{rno}}}$
}
Step 3. Solve $\mathbf{V}_1^{(1)}$ from (4.12).
```

Algorithm 1

During the decomposition of (4.14), the right-hand side of the right child is split after updating it recursively as follows:

$$\mathbf{T}_k^{(r)} - 4\mathbf{A}_E\mathbf{V}_k^{(r-1)}\mathbf{1}_{m/2^k} = \left[\mathbf{T}_{k+1}^{(2r-1)} \quad \mathbf{T}_{k+1}^{(2r)}\right]. \tag{4.23}$$

This splitting and updating sequence is a preorder traversal of the perfect binary tree from root node ②. The unknown matrix $\mathbf{V}_1^{(2)}$ is obtained by merging all column vectors $\mathbf{V}_J^{(r)}$ $(r = 2^{J-1} + 1, \ldots, 2^J)$. This sequence is a postorder traversal of the perfect binary tree from root node ②. To update (4.23), we need $\mathbf{V}_k^{(r-1)}$ which is obtained from the left child. Hence, to solve (4.22), it is necessary to update, split, and solve by using the following combined preorder and postorder traversal method.

The pseudocode of the proposed algorithm is as in Algorithm 1.

For example, at resolution scale $J = 4$, the proposed combined preorder and postorder traversal method is illustrated in Figure 2.

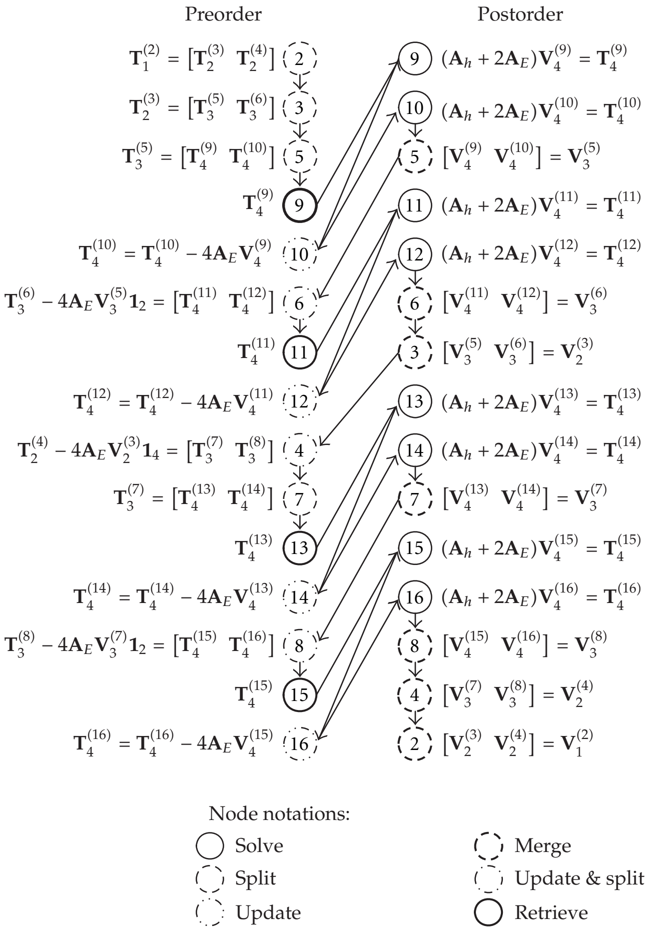

Preorder
Postorder
$\mathbf{T}_1^{(2)} = [\mathbf{T}_2^{(3)}\ \mathbf{T}_2^{(4)}]$ 2
$\mathbf{T}_2^{(3)} = [\mathbf{T}_3^{(5)}\ \mathbf{T}_3^{(6)}]$ 3
$\mathbf{T}_3^{(5)} = [\mathbf{T}_4^{(9)}\ \mathbf{T}_4^{(10)}]$ 5
$\mathbf{T}_4^{(9)}$ 9
$\mathbf{T}_4^{(10)} = \mathbf{T}_4^{(10)} - 4\mathbf{A}_E\mathbf{V}_4^{(9)}$ 10
$\mathbf{T}_3^{(6)} - 4\mathbf{A}_E\mathbf{V}_3^{(5)}\mathbf{1}_2 = [\mathbf{T}_4^{(11)}\ \mathbf{T}_4^{(12)}]$ 6
$\mathbf{T}_4^{(11)}$ 11
$\mathbf{T}_4^{(12)} = \mathbf{T}_4^{(12)} - 4\mathbf{A}_E\mathbf{V}_4^{(11)}$ 12
$\mathbf{T}_2^{(4)} - 4\mathbf{A}_E\mathbf{V}_2^{(3)}\mathbf{1}_4 = [\mathbf{T}_3^{(7)}\ \mathbf{T}_3^{(8)}]$ 4
$\mathbf{T}_3^{(7)} = [\mathbf{T}_4^{(13)}\ \mathbf{T}_4^{(14)}]$ 7
$\mathbf{T}_4^{(13)}$ 13
$\mathbf{T}_4^{(14)} = \mathbf{T}_4^{(14)} - 4\mathbf{A}_E\mathbf{V}_4^{(13)}$ 14
$\mathbf{T}_3^{(8)} - 4\mathbf{A}_E\mathbf{V}_3^{(7)}\mathbf{1}_2 = [\mathbf{T}_4^{(15)}\ \mathbf{T}_4^{(16)}]$ 8
$\mathbf{T}_4^{(15)}$ 15
$\mathbf{T}_4^{(16)} = \mathbf{T}_4^{(16)} - 4\mathbf{A}_E\mathbf{V}_4^{(15)}$ 16
9 $(\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_4^{(9)} = \mathbf{T}_4^{(9)}$
10 $(\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_4^{(10)} = \mathbf{T}_4^{(10)}$
5 $[\mathbf{V}_4^{(9)}\ \mathbf{V}_4^{(10)}] = \mathbf{V}_3^{(5)}$
11 $(\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_4^{(11)} = \mathbf{T}_4^{(11)}$
12 $(\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_4^{(12)} = \mathbf{T}_4^{(12)}$
6 $[\mathbf{V}_4^{(11)}\ \mathbf{V}_4^{(12)}] = \mathbf{V}_3^{(6)}$
3 $[\mathbf{V}_3^{(5)}\ \mathbf{V}_3^{(6)}] = \mathbf{V}_2^{(3)}$
13 $(\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_4^{(13)} = \mathbf{T}_4^{(13)}$
14 $(\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_4^{(14)} = \mathbf{T}_4^{(14)}$
7 $[\mathbf{V}_4^{(13)}\ \mathbf{V}_4^{(14)}] = \mathbf{V}_3^{(7)}$
15 $(\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_4^{(15)} = \mathbf{T}_4^{(15)}$
16 $(\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_4^{(16)} = \mathbf{T}_4^{(16)}$
8 $[\mathbf{V}_4^{(15)}\ \mathbf{V}_4^{(16)}] = \mathbf{V}_3^{(8)}$
4 $[\mathbf{V}_3^{(7)}\ \mathbf{V}_3^{(8)}] = \mathbf{V}_2^{(4)}$
2 $[\mathbf{V}_2^{(3)}\ \mathbf{V}_2^{(4)}] = \mathbf{V}_1^{(2)}$
Node notations:
Solve
Split
Update
Merge
Update & split
Retrieve

Figure 2: The combined preorder and postorder traversal for resolution scale $J = 4$.

In Figure 2, nodes 2, 3, 5, and 9 of preorder traversal are done at Step 1 and the remaining nodes are processed at Step 2. The computational efficiency of the proposed method is discussed in the next section.

5. An illustrated example

In this section, an example is presented to illustrate the proposed algorithm. We consider a singular linear system of (4.1) with

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} -33 & 0 & 1.0 & 0 \\ 0 & 1 & 0 & 1.0 \\ 0 & 621.4 & -28.27 & 0 \\ 0 & -327.1 & 12.72 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 52.65 \\ -23.69 \end{bmatrix}, \tag{5.1}$$

and $\mathbf{X}_0 = \begin{bmatrix} 0 & 0.5 & 1.0 & 0 \end{bmatrix}^{\mathrm{T}}$. And we assume that $u(t)$ is a unit step function. In the cases of $J = 4$ and 8, the simulation results are depicted in Figures 3 and 4, respectively.

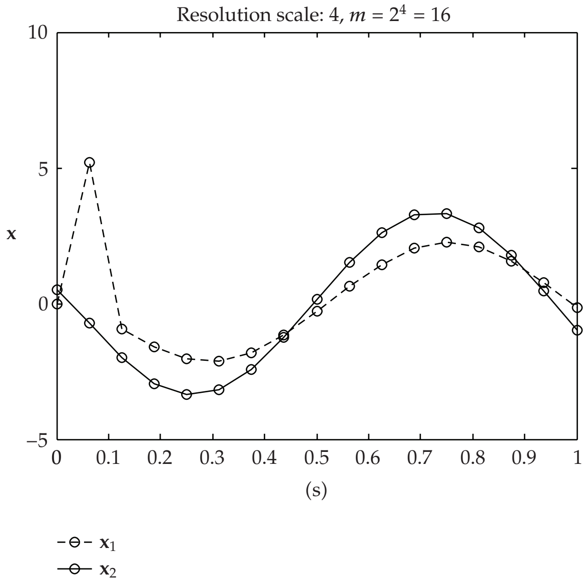


Figure 3: Case for resolution scale $J = 4$.

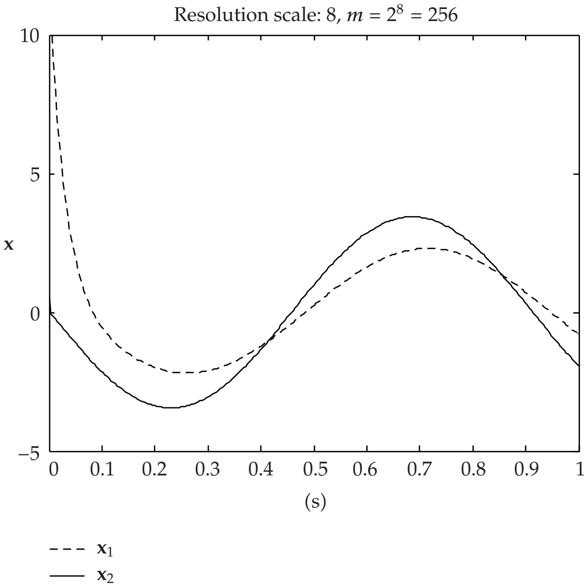


Figure 4: Case for resolution scale $J = 8$.

From these figures, it is clear that the solution accuracy is improved when the resolution scale is increased. However, it requires more computational time.

In (4.7), the LU factorization of $\mathbf{I}_m \otimes \mathbf{E} + \mathbf{P}_m^{\mathrm{T}} \otimes \mathbf{A}$ involves $O(m^3p^3)$ flops. The cost of the proposed algorithm is the sum of the cost of *WinSolver*, $O((m/2)p^3 + (m/2)p^2)$, and the cost of *WinTree*, $O(\sum_{k=1}^{J-2}(2^{J-1}p^2 + (2^{J+2k-2} - 2^{J-2})p))$ (see Appendix B). Since $m = 2^J$, the costs

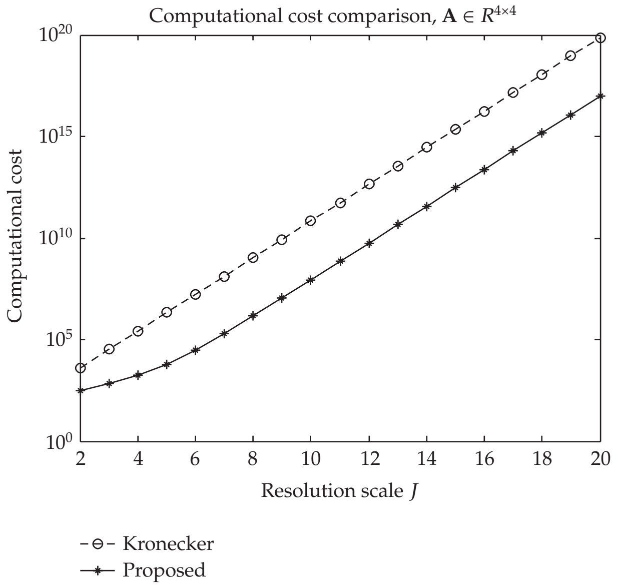


Figure 5: Log plot of flop counts for the Kronecker product method and the proposed method.

Table 1: Flop counts for various sizes of the matrix **A** and resolution scales.

	$\mathbf{A} \in R^{4\times 4}$				$\mathbf{A} \in R^{20\times 20}$			
J	Kronecker	Proposed algorithm			Kronecker	Proposed algorithm		
	method	*WinSolver*	*WinTree*	Total	method	*WinSolver*	*WinTree*	Total
2	4096	160	0	160	512000	16800	0	16800
5	2097152	1280	3360	4640	262144000	134400	32160	166560
10	6.871×10^{10}	40960	89534464	89575424	8.589×10^{12}	4300800	448983040	453283840
12	4.398×10^{12}	163840	5.726×10^{9}	5.727×10^{9}	5.497×10^{14}	17203200	2.864×10^{10}	2.867×10^{10}
15	2.251×10^{15}	1310720	2.932×10^{12}	2.932×10^{12}	2.814×10^{17}	137625600	1.466×10^{13}	1.466×10^{13}
18	1.152×10^{18}	10485760	1.501×10^{15}	1.501×10^{15}	1.441×10^{20}	1101004800	7.506×10^{15}	7.506×10^{15}
20	7.378×10^{19}	83886080	4.194×10^{16}	4.194×10^{16}	9.223×10^{21}	4.404×10^{9}	4.803×10^{17}	4.803×10^{17}

of *WinSolver* and $O(m^3p^3)$ can be rewritten as $O(2^{J-1}p^3 + 2^{J-1}p^2)$ and $O(2^{3J}p^3)$, respectively. Thus, the total cost of the proposed algorithm is

$$O\left(2^{J-1}p^3 + 2^{J-1}p^2 + \sum_{k=1}^{J-2}\left(2^{J-1}p^2 + \left(2^{J+2k-2} - 2^{J-2}\right)p\right)\right) \text{ flops.} \tag{5.2}$$

Table 1 and Figure 5 show that the computational cost of the proposed algorithm is significantly less than the Kronecker product method, and that the flop counts are increasing rapidly with resolution scale. As the resolution scale grows, the flop counts of *WinTree* is increasing more rapidly than that of *WinSolver* since the sizes of matrices $\mathbf{1}_m$, $\mathbf{T}_m$, and $\mathbf{V}_m$ increase exponentially.

Table 2

Level	Size of $\mathbf{1}_{m/2^{J-\mathrm{rno}}}$	Size of $\mathbf{T}_{J-\mathrm{rno}}^{(r/2+1)}$ and $\mathbf{V}_{J-\mathrm{rno}}^{(r/2)}$	Times	Computational cost
2	$2^{J-2} \times 2^{J-2}$	$p \times 2^{J-2}$	1	$p2^{2J-3} + (p^2+p)2^{J-1} - p2^{J-1}$
3	$2^{J-3} \times 2^{J-3}$	$p \times 2^{J-3}$	2	$p2^{2J-5} + (p^2+p)2^{J-2} - p2^{J-2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$J-k$	$2^k \times 2^k$	$p \times 2^k$	2^{J-k-2}	$2^{k+1}p^2 + 2^k(2^{2k}-1)p \times 2^{J-k-2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$J-2$	4×4	$p \times 4$	2^{J-4}	$2^3p^2 + 2^2(2^4-1)p \times 2^{J-4}$
$J-1$	2×2	$p \times 2$	2^{J-3}	$4p^2 + 2^1(2^2-1)p \times 2^{J-3}$

6. Conclusions

An efficient computational method was presented for state space analysis of singular systems via Haar wavelets. The problem was formulated as a generalized Sylvester matrix equation. We presented an explicit expression for the inverse of the Haar matrix and a combined preorder and postorder traversal algorithm to solve the problem more effectively. The full-order generalized Sylvester matrix equation was solved in terms of the solutions of simple linear matrix equations by the proposed algorithm. The efficiency of the proposed method was demonstrated by a numerical example.

Appendices

A. Formula for $\mathbf{C}_m$

In this appendix, we derive a formula for $\mathbf{C}_m$. By using (3.13), (3.9), and (3.10), we can write

$$\begin{aligned}
\mathbf{C}_m &= \mathbf{H}_m^{-1}\mathbf{P}_m\mathbf{H}_m \\
&= \begin{bmatrix} \mathbf{H}_{m/2}^{-1} \otimes \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} & \mathbf{I}_{m/2} \otimes \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} \end{bmatrix} \begin{bmatrix} \mathbf{P}_{m/2} & -\dfrac{1}{2m}\mathbf{H}_{m/2} \\ \dfrac{1}{2m}\mathbf{H}_{m/2}^{-1} & \mathbf{0}_{m/2} \end{bmatrix} \begin{bmatrix} \mathbf{H}_{m/2} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix} \\ \mathbf{I}_{m/2} \otimes \begin{bmatrix} 1 & -1 \end{bmatrix} \end{bmatrix} \\
&= \left[\left(\mathbf{H}_{m/2}^{-1} \otimes \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \right) \begin{bmatrix} \mathbf{P}_{m/2} & -\dfrac{1}{2m}\mathbf{H}_{m/2} \end{bmatrix} + \left(\mathbf{I}_{m/2} \otimes \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} \right) \begin{bmatrix} \dfrac{1}{2m}\mathbf{H}_{m/2}^{-1} & \mathbf{0}_{m/2} \end{bmatrix} \right] \\
&\quad \times \begin{bmatrix} \mathbf{H}_{m/2} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix} \\ \mathbf{I}_{m/2} \otimes \begin{bmatrix} 1 & -1 \end{bmatrix} \end{bmatrix} \\
&= \left(\left[\left(\mathbf{H}_{m/2}^{-1} \otimes \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \right) \mathbf{P}_{m/2} \quad -\frac{1}{2m} \left(\mathbf{H}_{m/2}^{-1} \otimes \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \right) \mathbf{H}_{m/2} \right] \right. \\
&\qquad \left. + \left[\frac{1}{2m}\mathbf{I}_{m/2} \otimes \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} \mathbf{H}_{m/2}^{-1} \quad \mathbf{0}_{m/2} \right] \right) \begin{bmatrix} \mathbf{H}_{m/2} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix} \\ \mathbf{I}_{m/2} \otimes \begin{bmatrix} 1 & -1 \end{bmatrix} \end{bmatrix} \\
&= \left[\left(\mathbf{H}_{m/2}^{-1} \otimes \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \right) \mathbf{P}_{m/2} \quad -\frac{1}{2m} \left(\mathbf{H}_{m/2}^{-1} \otimes \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \right) \mathbf{H}_{m/2} \right] \begin{bmatrix} \mathbf{H}_{m/2} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix} \\ \mathbf{I}_{m/2} \otimes \begin{bmatrix} 1 & -1 \end{bmatrix} \end{bmatrix} \\
&\quad + \left[\frac{1}{2m}\mathbf{I}_{m/2} \otimes \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} \mathbf{H}_{m/2}^{-1} \quad \mathbf{0}_{m/2} \right] \begin{bmatrix} \mathbf{H}_{m/2} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix} \\ \mathbf{I}_{m/2} \otimes \begin{bmatrix} 1 & -1 \end{bmatrix} \end{bmatrix}
\end{aligned}$$

$$= \left(\mathbf{H}_{m/2}^{-1} \otimes \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}\right) \mathbf{P}_{m/2}\left(\mathbf{H}_{m/2} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix}\right) - \frac{1}{2m}\left(\mathbf{H}_{m/2}^{-1} \otimes \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}\right) \mathbf{H}_{m/2}\left(\mathbf{I}_{m/2} \otimes \begin{bmatrix} 1 & -1 \end{bmatrix}\right)$$
$$+ \frac{1}{2m}\mathbf{I}_{m/2} \otimes \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} \mathbf{H}_{m/2}^{-1}\left(\mathbf{H}_{m/2} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix}\right). \tag{A.1}$$

Since $(\mathbf{A} \otimes \mathbf{B})\mathbf{C} = (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes 1) = (\mathbf{AC} \otimes \mathbf{B})$ and $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})$, the above equation is rewritten as

$$\begin{aligned}
\mathbf{C}_m &= \left((\mathbf{H}_{m/2}^{-1}\mathbf{P}_{m/2}) \otimes \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}\right)\left(\mathbf{H}_{m/2} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix}\right) - \frac{1}{2m}\left((\mathbf{H}_{m/2}^{-1}\mathbf{H}_{m/2}) \otimes \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}\right)\left(\mathbf{I}_{m/2} \otimes \begin{bmatrix} 1 & -1 \end{bmatrix}\right) \\
&\quad + \frac{1}{2m}\left((\mathbf{I}_{m/2}\mathbf{H}_{m/2}^{-1}) \otimes \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}\right)\left(\mathbf{H}_{m/2} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix}\right) \\
&= (\mathbf{H}_{m/2}^{-1}\mathbf{P}_{m/2}\mathbf{H}_{m/2}) \otimes \left(\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}\begin{bmatrix} 1 & 1 \end{bmatrix}\right) - \frac{1}{2m}(\mathbf{I}_{m/2}) \otimes \left(\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}\begin{bmatrix} 1 & -1 \end{bmatrix}\right) \\
&\quad + \frac{1}{2m}(\mathbf{I}_{m/2}) \otimes \left(\begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}\begin{bmatrix} 1 & 1 \end{bmatrix}\right) \\
&= \mathbf{C}_{m/2} \otimes \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} + \frac{1}{2m}\mathbf{I}_{m/2} \otimes \left(\begin{bmatrix} -0.5 & 0.5 \\ -0.5 & 0.5 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.5 \\ -0.5 & -0.5 \end{bmatrix}\right) \\
&= \mathbf{C}_{m/2} \otimes \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} + \frac{1}{2m}\mathbf{I}_{m/2} \otimes \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{2}\mathbf{C}_{m/2} & \frac{1}{m}\mathbf{1}_{m/2} \\ \mathbf{0}_{m/2} & \frac{1}{2}\mathbf{C}_{m/2} \end{bmatrix}.
\end{aligned} \tag{A.2}$$

B. Flop counts of the combined preorder and postorder traversal algorithm

In this appendix, we show that the computational cost for the combined preorder and postorder traversal algorithm described in Section 4.2 can be obtained as follows:

(1) WinSolve

Solve for $\mathbf{V}_J^{(r)}$ the system $(\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_J^{(r)} = \mathbf{T}_J^{(r)}$: $O(p^3)$.

Update $\mathbf{T}_J^{(r+1)}$ according to $\mathbf{T}_J^{(r+1)} = \mathbf{T}_J^{(r+1)} - 4\mathbf{A}_E\mathbf{V}_J^{(r)}$: $O(p(2p-1)+p) = O(2p^2)$.

Solve for $\mathbf{V}_J^{(r+1)}$ the system $(\mathbf{A}_h + 2\mathbf{A}_E)\mathbf{V}_J^{(r+1)} = \mathbf{T}_J^{(r+1)}$: $O(p^3)$.

The total iteration number of "for $(r = 2^{J-1} + 1;\ r < 2^J;\ r = r + 2)$" is $m/4$. Thus, *WinSolve* involves $O((m/4)(p^3+2p^2 + p^3)) = O((m/2)(p^3+\mathrm{p}^2))$ flops.

(2) WinTree

`Update and split` $\mathbf{T}_{J-\mathrm{rno}}^{(r/2+1)} - 4\mathbf{A}_E \mathbf{V}_{J-\mathrm{rno}}^{(r/2)} \mathbf{1}_{m/2^{J-\mathrm{rno}}}$ (see Table 2).

Therefore, the computational cost for *WinTree* can be calculated by

$$O\left(\sum_{k=1}^{J-2} \left(2^{k+1}p^2 + 2^k\left(2^{2k} - 1\right)p\right) \times 2^{J-k-2}\right) = O\left(\sum_{k=1}^{J-2} \left(2^{J-1}p^2 + \left(2^{J+2k-2} - 2^{J-2}\right)p\right)\right). \tag{A.1}$$

References

[1] C. Cattani, "Haar wavelet-based technique for sharp jumps classification," *Mathematical and Computer Modelling*, vol. 39, no. 2-3, pp. 255–278, 2004.

[2] C. Cattani, "Wavelet approach to stability-of-orbits analysis," *International Applied Mechanics*, vol. 42, no. 6, pp. 721–727, 2006.

[3] C. F. Chen and C. H. Hsiao, "Haar wavelet method for solving lumped and distributed-parameter systems," *IEE Proceedings: Control Theory and Applications*, vol. 144, no. 1, pp. 87–94, 1997.

[4] C. F. Chen and C.-H. Hsiao, "Wavelet approach to optimising dynamic systems," *IEE Proceedings: Control Theory and Applications*, vol. 146, no. 2, pp. 213–219, 1999.

[5] F. L. Lewis, "A survey of linear singular systems," *Circuits, Systems, and Signal Processing*, vol. 5, no. 1, pp. 3–36, 1986.

[6] F. L. Lewis and B. G. Mertzios, "Analysis of singular systems using orthogonal functions," *IEEE Transactions on Automatic Control*, vol. 32, no. 6, pp. 527–530, 1987.

[7] R. Kalpana and S. R. Balachandar, "Haar wavelet method for the analysis of transistor circuits," *AEU - International Journal of Electronics and Communications*, vol. 61, no. 9, pp. 589–594, 2007.

[8] C. F. van Loan, "The ubiquitous Kronecker product," *Journal of Computational and Applied Mathematics*, vol. 123, no. 1-2, pp. 85–100, 2000.

[9] A. Haar, "Zur Theorie der orthogonalen Funktionensysteme," *Mathematische Annalen*, vol. 69, no. 3, pp. 331–371, 1910.

[10] Z. Gajic and M. T. J. Qureshi, *Lyapunov Matrix Equation in System Stability and Control*, vol. 195 of *Mathematics in Science and Engineering*, Academic Press, San Diego, Calif, USA, 1995.

[11] B.-S. Kim, I.-J. Shim, B. K. Choi, and J. H. Jeong, "Wavelet based control for linear systems via reduced order Sylvester equation," in *Proceedings of the 3rd International Conference on Cooling and Heating Technologies (ICCHT '07)*, pp. 239–244, Tokyo, Japan, July 2007.

[12] F. Carrano and W. Savitch, *Data Structures and Abstractions with Java*, Prentice Hall, Upper Saddle River, NJ, USA, 2003.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 687318, 7 pages
doi:10.1155/2008/687318

Research Article

On the Discrete Harmonic Wavelet Transform

Carlo Cattani[1] **and Aleksey Kudreyko**[2]

[1] *Department of Pharmaceutical Sciences (DiFarma), University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy*
[2] *Department of Mathematics and Computer Science, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy*

Correspondence should be addressed to Aleksey Kudreyko, skateswoosh84@yahoo.com

Received 29 May 2008; Accepted 26 July 2008

Recommended by Cristian Toma

The discrete harmonic wavelet transform has been reviewed and applied towards given functions. The absolute error of reconstruction of the functions has been computed.

1. Introduction

The discrete harmonic wavelet transform was developed by Newland in 1993 [1, 2]. Similar to the ordinary discrete wavelet transform, the classical harmonic wavelet transform can also perform multiresolution analysis of a function. In addition, it has a fast algorithm based on fast Fourier transform for numerical implementation. A distinct advantage of harmonic wavelets is that they are disjoint in frequency domain (see Figure 1) and the Fourier transform of the successive levels decreases in propagation of their bandwidth (1.1).

$$\widehat{\varphi}(\omega)=\begin{cases}\left(\dfrac{1}{2\pi}\right)\left(\dfrac{1}{2^j}\right) & \text{for } 2\pi 2^j \le \omega < 4\pi 2^j,\\ 0 & \text{elsewhere.}\end{cases} \tag{1.1}$$

Calculating its inverse Fourier transform, we obtain

$$\varphi_k^j(x)=\frac{e^{4\pi i(2^j x-k)}-e^{2\pi i(2^j x-k)}}{2\pi i(2^j x-k)}, \tag{1.2}$$

where $j=0,\dots,\infty$ and $k=-\infty,\dots,\infty$. This function represents a class of pulsed functions due to its compact support in the space domain.

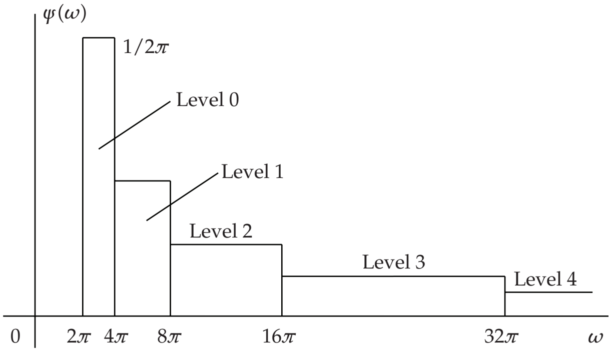


Figure 1: Values of the Fourier transform of harmonic wavelets of different levels.

2. Discretisation of a real function

The goal of the wavelet transform is to decompose any arbitrary given function $f(x)$ into an infinite summation of wavelets at different scales according to the expansion

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} a_{j,k} \psi_k^j(x), \tag{2.1}$$

or in the alternative form [3]

$$f(x) = \sum_{k=-\infty}^{\infty} a_{\phi,k} \phi(x-k) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} a_{j,k} \psi_k^j(x). \tag{2.2}$$

The first sum is a smooth approximation of $f(x)$, where the wavelets for $j \leq 0$ have been rolled together into scaling functions. The second sum is an addition of the details of $f(x)$ at a specific level of resolution.

For complex wavelet coefficients, we have to define two amplitude coefficients

$$a_{j,k}= 2^j \int_{-\infty}^{\infty} f(x) \psi^*(2^j x - k) dx, \qquad \tilde{a}_{j,k}= 2^j \int_{-\infty}^{\infty} f(x) \psi(2^j x - k) dx, \tag{2.3}$$

and the corresponding pair of complex coefficients for the terms of scaling function,

$$a_{\varphi,k} = \int_{-\infty}^{\infty} f(x) \varphi^*(x-k) dx, \qquad \tilde{a}_{\varphi,k} = \int_{-\infty}^{\infty} f(x) \varphi(x-k) dx. \tag{2.4}$$

If $f(x)$ is real, then $\tilde{a}_{j,k}$ is the complex conjugate of $a_{j,k}$, that is, $\tilde{a}_{j,k} = a_{j,k}^*$, but to allow the general case, when $f(x)$ is complex, we will consider $\tilde{a}_{j,k}$ and $a_{j,k}^*$ as two different amplitudes. Then the expansion formulas (2.1) and (2.2) become [2]

$$\begin{aligned} f(x) &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \{ a_{j,k} \psi(2^j x - k) + \tilde{a}_{j,k} \psi^*(2^j x - k) \}, \\ f(x) &= \sum_{k=-\infty}^{\infty} \{ a_{\varphi,k} \varphi(x-k) + \tilde{a}_{\varphi,k} \varphi^*(x-k) \} \\ &\quad + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} \{ a_{j,k} \psi(2^j x - k) + \tilde{a}_{j,k} \psi^*(2^j x - k) \}. \end{aligned} \tag{2.5}$$

Our primary purpose is to compute the coefficients $a_{\varphi,k}$, $\widetilde{a}_{\varphi,k}$, $a_{j,k}$ and $\widetilde{a}_{j,k}$ of this expansion.

An important condition for the function is that

$$\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty. \tag{2.6}$$

Let us consider a real-valued function $f(x)$, represented by its discrete sequence

$$f_r, \quad r = 0, 1, \ldots, N-1, \tag{2.7}$$

where $N= 2^j$. Recalling the definition of the discrete Fourier transform, the corresponding Fourier coefficients are

$$\widehat{f}_m = \frac{1}{N}\sum_{r=0}^{N-1} f_r e^{-2\pi i m r/N}, \quad m = 0, 1, \ldots, N-1. \tag{2.8}$$

Note that

$$\widehat{f}_{N-m} = \frac{1}{N}\sum_{r=0}^{N-1} f_r e^{-2\pi i (N-m) r/N} = \frac{1}{N}\sum_{r=0}^{N-1} f_r e^{-2\pi i r} e^{2\pi i m r/N} = \widehat{f}^*_m, \tag{2.9}$$

where the asterisk stands for the complex conjugate; $\widehat{f}_0$ and $\widehat{f}_{N/2}$ are always real numbers.

Furthermore, we will consider the coefficient $a_{j,k}$, defined by the first formula in (2.3). Firstly, we will substitute $\psi^*_{j,k}(x)$ in terms of its Fourier transform (1.1)

$$\psi^*_{j,k}(x) = \frac{1}{2^j}\int_{2\pi 2^j}^{4\pi 2^j} \frac{1}{2\pi} e^{i\omega k/2^j} e^{-i\omega x} d\omega \tag{2.10}$$

into the first formula of (2.3), and we obtain the following integral

$$a_{j,k} = \frac{1}{2\pi}\int_{2\pi 2^j}^{4\pi 2^j} e^{i\omega k/2^j} d\omega \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx, \tag{2.11}$$

where we have reversed the order of integration. The second integral over x represents the Fourier transform of $f(x)$ multiplied by 2π, and (2.11) becomes

$$a_{j,k} = \int_{2\pi 2^j}^{4\pi 2^j} \widehat{f}(\omega) e^{-i\omega k/2^j} d\omega. \tag{2.12}$$

To derive a discrete algorithm of decomposition of the function, we must replace the operation of integration by summation, and (2.12) becomes

$$a_{2^j+k} = \sum_{s=0}^{2^j-1} \widehat{f}_{2^j+s} e^{2\pi i s k/2^j}, \quad k = 0, \ldots, 2^j - 1. \tag{2.13}$$

This identity represents the inverse discrete Fourier transform for the sequence of frequency coefficients $\widehat{f}_{2^j+s}$.

Analogous transformation towards the computation of $\tilde{a}_{2^j+k}$ will lead us to the following [2]:

$$\tilde{a}_{2^j+k} = \sum_{s=0}^{2^j-1} \hat{f}_{N-(2^j+s)} e^{2\pi isk/2^j}, \quad k = 0, \dots, 2^j - 1. \tag{2.14}$$

Computation of the amplitudes a_0 and $a_{N/2}$ in the reviewed algorithm involves special approach, and $a_0 = \hat{f}_0$ and $a_{N/2} = \hat{f}_{N/2}$ [2].

Also, it is easy to show from (2.13) that if $j = 0$, then $k = 0$ and

$$a_1 = \hat{f}_1. \tag{2.15}$$

Summarizing the stated above, the sequence of operations for computation of wavelet amplitude coefficients is as follows:

(i) represent the given function $f(x)$ by a discrete sequence f_r, where $r = 0, 1, \dots, N-1$;

(ii) compute the set of frequency coefficients by fast Fourier transform $\hat{f}_m$, where $m = 0, 1, \dots, N-1$;

(iii) the inverse fast Fourier transform of the octave blocks $\hat{f}_m$ generates the amplitudes of the harmonic wavelet expansion of the function f_r.

It is important to mention that this algorithm works for only the functions which satisfy the following conditions.

(i) The discrete transform covers the unit internal of x.

(ii) The analysed function is periodic in x with period 1.

The algorithm was applied to the given functions which satisfy the mentioned conditions.

3. Implementation of Newland's algorithm towards a given function

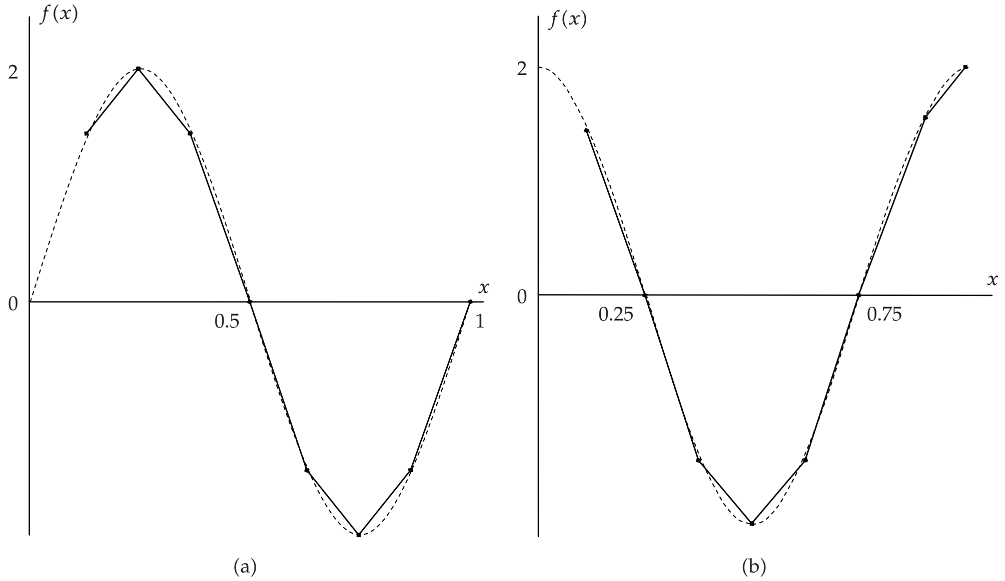
Let us review functions which satisfy the stated conditions. For example, it is $f(x) = 2\sin 2\pi x$ and $f(x) = 2\cos 2\pi x$. Following the algorithm, we discretise the interval $[0;1]$ into $N = 2^j$ equally spaced nods, and obtain discrete set of values of functions

$$f_r = 2\sin\frac{2\pi r}{N}, \quad f_r = 2\cos\frac{2\pi r}{N}, \quad r = 0, \dots, N-1. \tag{3.1}$$

The fast Fourier transform (2.8) of the obtained discrete sequence gives us the set Fourier coefficients $\hat{f}_m$. Recalling that $a_0 = \hat{f}_0$, $a_1 = \hat{f}_1$, and $a_{N/2} = \hat{f}_{N/2}$, we can easily find these three coefficients. Another part of coefficients from a_{2^j} to $a_{2^{j+1}-1}$ is obtained by computation of the inverse fast s Fourier transform (2.13) of coefficients from $\hat{f}_{2^j}$ to $\hat{f}_{2^{j+1}-1}$.

To reconstruct the function from its wavelet coefficients, we followed the reverse algorithm of decomposition, that is: the fast Fourier transform of the wavelet coefficients a_{2^j+k} represents the discrete Fourier transform of the reconstructed function f_r. Then, taking into account the shifting property (2.9), we can find f as inverse fast Fourier transform of $\hat{f}$.

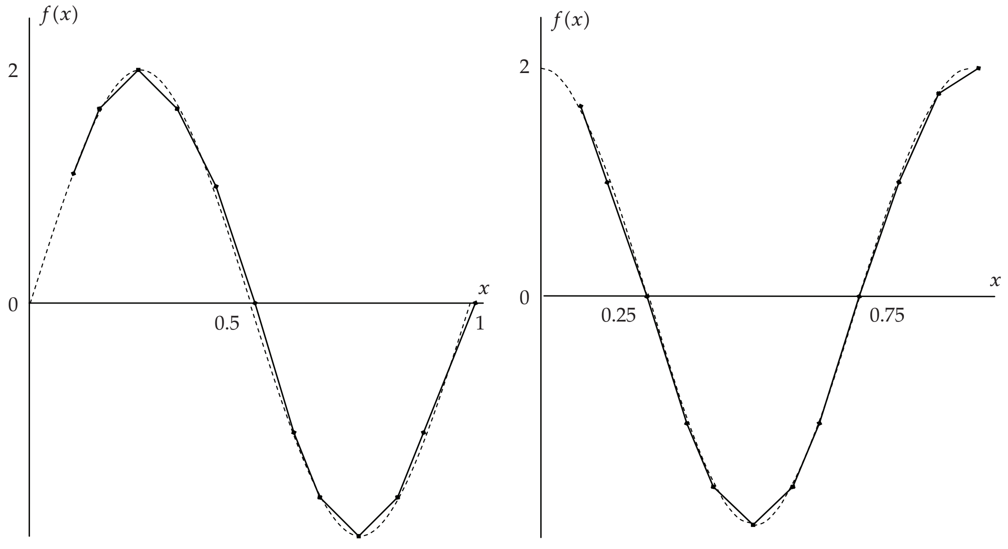
The results of decomposition and reconstruction of functions $f(x) = 2\sin 2\pi x$ and $f(x) = 2\cos 2\pi x$ are presented in Figures 2 and 3.



(a)

(b)

Figure 2: Arbitrary given function: (a) $\sin 2\pi x$, (b) $\cos 2\pi x$ (dashed line), and its reconstructed clone (solid line) from wavelet coefficients for $N = 8$.



(a)

(b)

Figure 3: Arbitrary given function: (a) $\sin 2\pi x$, (b) $\cos 2\pi x$ (dashed line), and its reconstructed clone (solid line) from wavelet coefficients for $N = 16$.

One can notice that the plots of the reconstructed functions are defined within the interval from $r = 1$ to $r = N$. The difference between the algorithm and its corresponding computer code consists in that we put$a_1$in the code instead of a_0, and so forth . Therefore, the reconstruction of the function begins from point $1/N$ to 1, and not from 0 to $N - 1$.

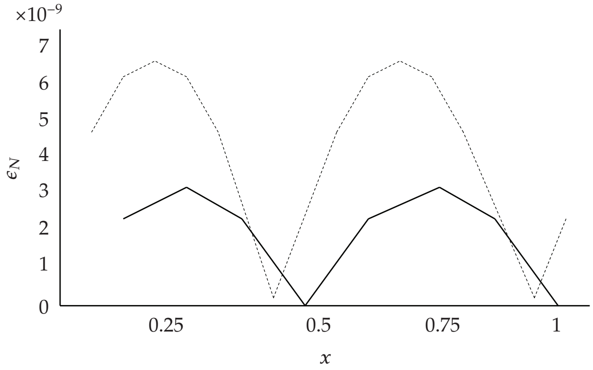


Figure 4: Absolute error of the reconstruction of $f(x) = 2\sin 2\pi x$ for $N = 8$ (solid line) and $N = 16$ (dashed line).

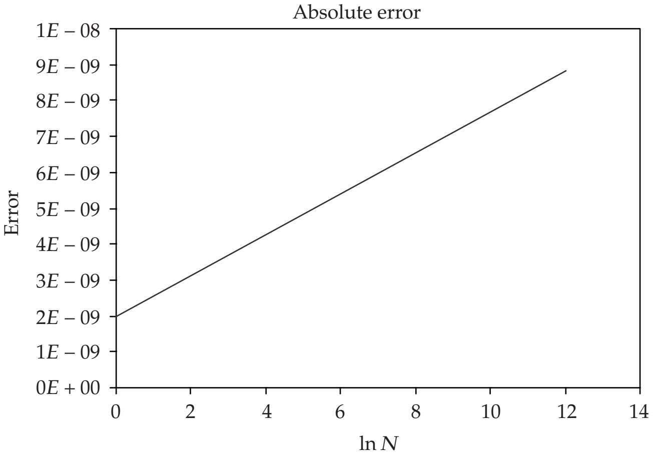


Figure 5: Absolute error of the reconstruction of $2\sin 2\pi x$ after regression analysis.

To show the efficiency of the algorithm, it is worth to estimate the absolute error of the reconstructed function in the discrete nods. It is well known that the absolute error is given by

$$\epsilon_N = |f(x_r) - f_{\text{rec}}(x_r)|, \quad r = 0, \dots, N-1, \tag{3.2}$$

where $f_{\text{rec}}(x_r)$ is the value of the reconstructed point. The dependence of absolute error of the reconstruction of the function from $\ln N$ is represented in Figure 5 and for two partial cases, when $N = 8$ and $N = 16$ can be found in Figure 4. As we can see, small numbers of the level of decomposition j give a very good approximation, when we reconstruct the function.

4. Discussion of results and conclusion

Wavelets are considered as a new powerful tool for time-frequency analysis of nonlinear phenomena. In our paper, we discussed the harmonic wavelet transform and applied its algorithm towards decomposition and reconstruction of functions with a unit period. This algorithm might be useful for the wavelet solution of partial differential equations, when it is reduced to a system of ordinary differential equations [4, 5]. The algorithm of the decomposition consists of fast Fourier transform of the given discredited vector function, in which approximation error is proportional to $\ln N$ and the corresponding approximation was obtained in our simulations (see Figure 5). It means that the increase of the length of N leads us to a slow, but steady increase of the approximation error. The line of the dependence of the error from N was obtained by implementing the method of least squares [6]. Note that the line of the plot takes discrete values due to the fact that N takes only integer values of 2^j.

The only disadvantage of harmonic wavelets is that its decay rate is relatively low (proportional to x^{-1}), therefore, its localisation is not precise. However, we have this disadvantage for the restricted Fourier transform of a harmonic wavelet of a specific level.

The application of harmonic wavelets towards particular problems is still new. The subject is developing very fast, however, there are still many questions remain unanswered. For example, what is the best choice of wavelet to use for a particular problem? How far does the harmonic wavelet transform computational simplicity compensate its slow decay rate in the x-domain? How it can be used for the solution of integrodifferential equations, and many others. This work is in progress.

Acknowledgment

The work of Aleksey Kudreyko has been supported by Istituto Nazionale di Alta Matematica Francesco Severi (Rome-IT) under Scholarship U 2007/000394, 02/07/2007.

References

[1] D. E. Newland, “Harmonic wavelet analysis,” *Proceedings of the Royal Society of London. Series A*, vol. 443, no. 1917, pp. 203–225, 1993.

[2] D. E. Newland, *An Introduction to Random Vibrations, Spectral & Wavelet Analysis*, John Wiley & Sons, New York, NY, USA, 3rd edition, 1993.

[3] C. Cattani and J. Rushchitsky, *Wavelet and Wave Analysis as Applied to Materials with Micro or Nanostructure*, vol. 74 of *Series on Advances in Mathematics for Applied Sciences*, World Scientific, Hackensack, NJ, USA, 2007.

[4] C. Cattani and A. Kudreyko, “Mutiscale analysis of the Fisher equation,” in *Proceedings of International Conference on Computational Science and Its Applications (ICCSA ’08)*, vol. 5072 of *Lecture Notes in Computer Science*, pp. 1171–1180, Springer, Perugia, Italy, June-July 2008.

[5] S. V. Muniandy and I. M. Moroz, “Galerkin modelling of the Burgers equation using harmonic wavelets,” *Physics Letters A*, vol. 235, no. 4, pp. 352–356, 1997.

[6] B. P. Demidovich, I. A. Maron, and E. Z. Shuvalova, *Chislennie Metody Analiza*, Lan’, Moscow, Russia, 2008.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 654820, 7 pages
doi:10.1155/2008/654820

Research Article

Resolution of First- and Second-Order Linear Differential Equations with Periodic Inputs by a Computer Algebra System

M. Legua,[1] I. Morales,[2] and L. M. Sánchez Ruiz[3]

[1] *Departamento de Matemática Aplicada, Centro Politécnico Superior (CPS), Universidad de Zaragoza, 50015 Zaragoza, Spain*
[2] *Departamento de Matemática Aplicada, Escuela Técnica Superior de Ingenieros Agrónomos (ETSIA), and (IUMPA), Universidad Politécnica de Valencia, 46022 Valencia, Spain*
[3] *Departamento de Matemática Aplicada, Escuela Técnica Superior de Ingeniería del Diseño (ETSID), and Instituto Universitario de Matemática Pura y Aplicada (IUMPA), Universidad Politécnica de Valencia, 46022 Valencia, Spain*

Correspondence should be addressed to L. M. Sánchez Ruiz, lmsr@mat.upv.es

Received 16 June 2008; Accepted 26 July 2008

Recommended by Carlo Cattani

In signal processing, a pulse means a rapid change in the amplitude of a signal from a baseline value to a higher or lower value, followed by a rapid return to the baseline value. A square wave function may be viewed as a pulse that repeats its occurrence periodically but the return to the baseline value takes some time to happen. When these periodic functions act as inputs in dynamic systems, the standard tool commonly used to solve the associated initial value problem (IVP) is Laplace transform and its inverse. We show how a computer algebra system may also provide the solution of these IVP straight forwardly by adequately introducing the periodic input.

1. Introduction

Linear differential equations $L[y(t)] = f(t)$, where f is a known input and L is the nth-order linear differential operator,

$$\begin{gathered} L = D^n + a_{n-1}(t)D^{n-1} + \cdots + a_1(t)D + a_0(t), \\ D^i = \frac{d^i}{dt^i}, \quad 1 \le i \le n, \end{gathered} \tag{1.1}$$

are easily solved when it has got constant coefficients a_i, the roots of the homogeneous associated equation are known, and the input is an adequate combination of exponential, cosine, sine, and polynomial functions.

A different situation arises when, even the coefficients being constants, the input is a periodic function which may fail to be continuous or may be formed by a sequence of pulses. In this case, two widely used methods to handle the problem are Fourier series and Laplace transforms [1, 2]. Both methods require to know their properties which on the other hand are highly rewarding since they can help to solve some partial differential equations, too.

The proliferation in the use of computer algebra systems has facilitated and fastened obtaining some of these solutions. In this note, we show how some of them, namely DERIVE [3], may directly provide the solution of first- and second-order differential equations in the presence of periodic inputs.

2. Generating periodic functions

Let us recall that a function f is called *periodic* with period $T > 0$, if for all x in the domain of the function $f(x+T) = f(x)$. Geometrically, this means that the graph of f repeats itself every T units. Periodic functions do appear in a number of real-life situations such as alternating currents, the motion of a pendulum, vibrations of a spring, and sound waves, just to mention a few of them.

Computer algebra systems allow an easy representation of periodic functions since they have usually got an implemented command which enables to find the remainder on an integer division of two real numbers. With this goal, assume that we have a given function f defined in $[a, b[, b > a$, and is to generate a $(b-a)$-periodic function, $\mathrm{ext}(f_{[a,b[})$ which repeats the values of f in successive intervals of length $b-a$. Then, if DERIVE is the handy program, it will generate $\mathrm{ext}(f_{[a,b[})$ by just substituting the variable t of $f(t)$ by $a + \mathrm{MOD}(t-a, b-a)$ [4]. MATHEMATICA [5] and MATLAB [6] programs enjoy similar capabilities by means of their corresponding commands (cf. [7]).

Example 2.1. Represent the π-periodic function f generated as a full-wave rectified sinusoid such that

$$f(t) = \sin t, \quad t \in [0, \pi[. \tag{2.1}$$

Solution 1. The DERIVE program enables to introduce f by writing

$$\mathrm{SUBST}\big(\mathrm{SIN}(t), t, \mathrm{MOD}(t, \mathrm{PI})\big). \tag{2.2}$$

Simplifying the above expression and plotting the generated function, we obtain the graph that appears in Figure 1.

A similar graph is obtained with MATHEMATICA by introducing Plot(Sin(Mod(x,Pi, 0)), $\{x, -5, 5\}$, PlotRange $\{-1, 2\}$), and with MATLAB by setting, L = "sin(mod(x, pi))," ezplot(L, [−5, 5]).

From now on, we will focus on DERIVE which has also got a useful CHI (a, x, b) function that is equally to 1 in $]a, b[$ and vanishes outside this interval.

Example 2.2. Represent the 2π-periodic function such that

$$f(t) = \begin{cases} 20, & t \in [0, \pi[, \\ -20, & t \in [\pi, 2\pi[. \end{cases} \tag{2.3}$$

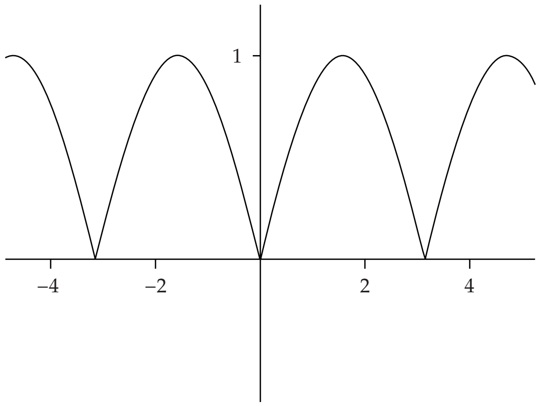


Figure 1: Full-wave rectified sinusoid.

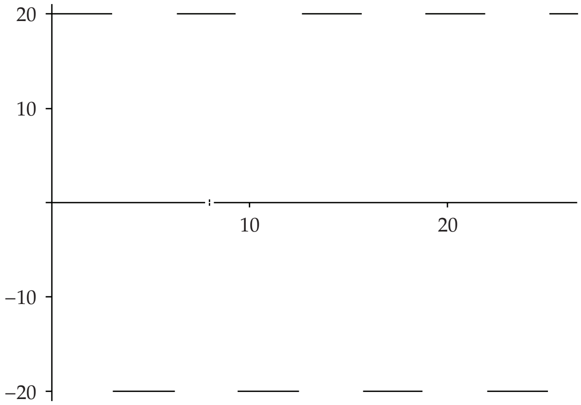


Figure 2: Square wave function.

Solution 2. Having in mind the above, we just have to introduce

$$\text{SUBST}\big(20\,\text{CHI}(0,t,\text{PI}) - 20\,\text{CHI}(\text{PI},t,2\text{PI}),t,\text{MOD}(t,2\text{PI})\big). \tag{2.4}$$

Simplifying the above expression and plotting the generated function, we obtain the graph that appears in Figure 2.

Finally and for the sake of completeness, let us recall the following result which justifies that $\text{ext}(f_{[a,b[})$ is a $(b-a)$-periodic function which coincides with f on $[a,b[$. Its easy proof follows immediately noting that if for each real number t, $I(t)$ denotes integer part function, that is, $I(t)$ equals the integer number n such that $n \leq t < n+1$, then $I((t-a)/(b-a)) = 0$ for $t \in [a,b[$.

Lemma 2.3. *If f is a real-valued function defined over $[a,b[$, then*

$$\text{ext}\Big(f_{[a,b[}\Big) = f\left(t - (b-a)I\left(\frac{t-a}{b-a}\right)\right) \tag{2.5}$$

is a $(b-a)$-periodic function defined over $\mathbb{R}$ that coincides with f in $[a,b[$.

3. Initial value problems with periodic inputs

Let us now recall some DERIVE commands that enable to solve first- and second-order linear differential equations. Given a linear differential equation written in the form

$$y' + p(x)y = q(x), \tag{3.1}$$

the command LINEAR1_GEN (p, q, x, y, c) provides the general solution in terms of the symbolic constant c. The command LINEAR1 (p, q, x, y, x_0, y_0) simplifies to the explicit solution for the initial condition $y = y_0$ at $x = x_0$, there being other available commands for other specific kinds of differential equations (cf. [8]).

Let us also recall that DSOLVE2 $(p, q, r, x, c1, c2)$ provides the general solution of the second-order linear differential equation

$$y'' + p(x)\cdot y' + q(x)\cdot y = r(x) \tag{3.2}$$

in terms of the symbolic constants c_1 and c_2. Analogously,

$$\text{DSOLVE2\_BV}(p, q, r, x, x_0, y_0, x_1, y_1) \tag{3.3}$$

is simplified to the explicit solution for the boundary value conditions $y = y_0$ at $x = x_0$, and $y = y_1$ at $x = x_1$, and DSOLVE2_IV$(p, q, r, x, x_0, y_0, v_0)$ to the explicit solution for the initial value conditions $y = y_0$ at $x = x_0$, and $y' = v_0$ at $x = x_0$.

Next, we provide an example of a first- (and another of a second-) order linear differential equation with periodic inputs and show how the aforementioned commands can cope with periodic inputs.

Example 3.1. Considering as input the function f of Example 2.2, solve

$$x' + x = f(t). \tag{3.4}$$

Solution 3. Let us combine the aforementioned implemented functions with the function f defined in Example 2.2 by

$$f(t) := \text{SUBST}\big(20\,\text{CHI}(0, t, \text{PI}) - 20\,\text{CHI}(\text{PI}, t, 2\text{PI}), t, \text{MOD}(t, 2\text{PI})\big). \tag{3.5}$$

The general integral is obtained by simplifying

$$\text{LINEAR1\_GEN}\big(1, f(t), t, x, c\big). \tag{3.6}$$

Hence, we obtain the solution (note that the following expression giving x is not corrupted. It is included exactly in the way provided by the computer algebra system since it has got

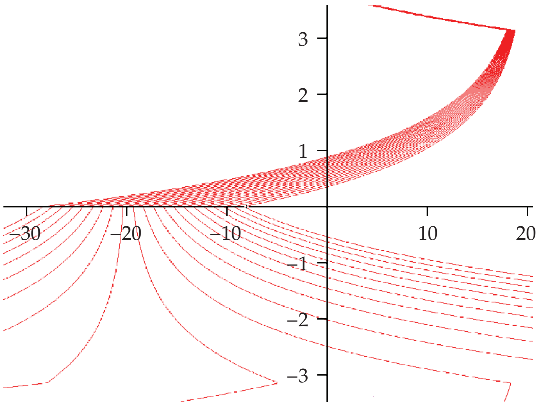


Figure 3: Particular solutions.

a long fraction and its terms are continuously written throughout different lines),

$$\mathtt{x} =$$

$$\frac{\hat{\mathrm{e}}^{-\mathrm{t}}\cdot\left(10(\hat{e}^{\pi}+1)\cdot \mathrm{SIGN}\left(2\pi\cdot \mathrm{FLOOR}\left(\frac{t}{2\pi}\right)-t+2\pi\right)\cdot(\hat{\mathrm{e}}^{2\text{ß}\cdot\mathrm{FLOOR}(\mathrm{t}/(2\text{ß}))+2\text{ß}}-\hat{\mathrm{e}}^{\mathrm{t}})+20\cdot(\hat{\mathrm{e}}^{\text{ß}}+1)\cdot\mathrm{SIGN}\left(2\text{ß}\cdot\mathrm{FLOOR}\left(\frac{t}{2\pi}\right)-\mathrm{t}+\text{ß}\right)\cdot(\hat{\mathrm{e}}^{\mathrm{t}}-\hat{\mathrm{e}}^{2\text{ß}\cdot\mathrm{FLOOR}(}\right.}{\hat{\mathrm{e}}^{\text{ß}}+}$$

$$\frac{\left.{}^{\mathrm{t}/(2\text{ß}))+\text{ß}})+10\cdot(\hat{\mathrm{e}}^{\text{ß}}+1)\cdot\mathrm{SIGN}\left(2\text{ß}\cdot\mathrm{FLOOR}\left(\frac{t}{2\pi}\right)-\mathrm{t}\right)\cdot(\hat{\mathrm{e}}^{2\text{ß}\cdot\mathrm{FLOOR}(\mathrm{t}/(2\text{ß}))}-\hat{\mathrm{e}}^{\mathrm{t}})-10\cdot\hat{\mathrm{e}}^{2\text{ß}\cdot\mathrm{FLOOR}(\mathrm{t}/(2\text{ß}))}\cdot(\hat{\mathrm{e}}^{3\text{ß}}-\hat{\mathrm{e}}^{2\text{ß}}+\hat{\mathrm{e}}^{\text{ß}}-1)+\mathrm{c}\cdot(\hat{\mathrm{e}}^{\text{ß}}+1)\right)}{1} \tag{3.7}$$

making c take a finite set of values, we get the corresponding particular solutions, *for example*, with the integer values between −10 and 10, we obtain 21 particular solutions whose graphs are depicted in Figure 3.

In what concerns second-order differential equations, let us recall that the *forced motion* of a mass m attached to a vibrating spring with damping constant α and spring constant k is modeled by

$$mx'' + \alpha x' + kx = f(t), \tag{3.8}$$

where $f(t)$ is an external force acting upon m. When the external force is identically equal to zero, the motion is called a *free motion* and it is well known that its solution (*underdamped*, *critically damped*, or *overdamped*) depends very heavily upon the nature of the characteristic roots. Next, we solve a problem where the external force is a wave square function (see [9, pages 336–341]).

Example 3.2. A mass of 1 g is attached to the end of a spring with $k = 20\,\text{dyn/cm}$ and the air resistance acts upon the mass with a force that is 4 times its velocity at time t. The mass has got no motion at its equilibrium position when it is subjected to an external periodic force equal to a square wave function of amplitude 20 cm. Find the position $x(t)$ of the mass.

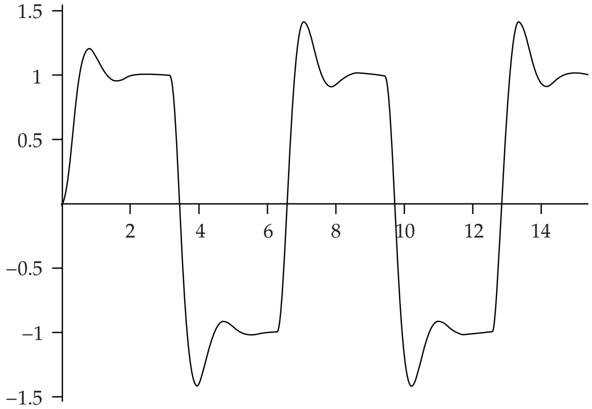


Figure 4: Position of the mass.

Solution 4. In this case, the position x is set by the initial value problem

$$x'' + 4x' + 20\,x = f(t), \qquad x(0) = x'(0) = 0, \tag{3.9}$$

where f is the 2π-periodic depicted in Example 2.2. Thus the position x is obtained by simplifying

$$\mathrm{DSOLVE2\_IV}(4, 20, f(t), t, 0, 0, 0). \tag{3.10}$$

The solution provided by the computer algebra system (remember that long fraction terms are continuously written throughout different lines) is the following one,

$$\frac{\hat{e}^{-2t}\cdot\Big((\hat{e}^{2\pi}+1)\cdot\mathrm{SIGN}\Big(2\pi\cdot\mathrm{FLOOR}\Big(\frac{t}{2\pi}\Big)-t+2\pi\Big)\cdot(\hat{e}^{4\text{ß}\cdot\mathrm{FLOOR}(t/(2\text{ß}))+4\text{ß}}\cdot(2\mathrm{COS}(4t)+\mathrm{SIN}(4t))-2\hat{e}^{2t})+2\cdot(\hat{e}^{2\text{ß}}+1)\cdot\mathrm{SIGN}\Big(2\text{ß}\cdot\mathrm{FLOOR}\Big(\frac{t}{2\pi}\Big) -t+\text{ß}\Big)\cdot(2\hat{e}^{2t}-\hat{e}^{4\text{ß}\cdot\mathrm{FLOOR}(t/(2\text{ß}))+2\text{ß}}\cdot(2\mathrm{COS}(4t)+\mathrm{SIN}(4t)))+(\hat{e}^{2\text{ß}}+1)\cdot\mathrm{SIGN}\Big(2\text{ß}\cdot\mathrm{FLOOR}\Big(\frac{t}{2\pi}\Big)-t\Big)\cdot(\hat{e}^{4\text{ß}\cdot\mathrm{FLOOR}(t/(2\text{ß}))}\cdot(2\mathrm{COS}(4t)+\mathrm{SIN}(4t))-2\hat{e}^{2t})+(1-\hat{e}^{2\text{ß}})\cdot(2\mathrm{COS}(4t)+\mathrm{SIN}(4t))\cdot(\hat{e}^{4\text{ß}\cdot\mathrm{FLOOR}(t/(2\text{ß}))}\cdot(\hat{e}^{4\text{ß}}+1)-2)\Big)}{4(\hat{e}^{2\text{ß}}+1)} \tag{3.11}$$

its graph is plotted in Figure 4.

4. Conclusion

Laplace transform is an important tool classically used to solve initial value problems in the presence of a periodic external force. For its adequate use, properties of direct and inverse Laplace transforms are required and they are extensively studied in many textbooks.

In this paper, we have seen how DERIVE enables to avoid this in a very simple way. This is achieved by using its standard routines that provide the general solution of a differential equation and the exact solution of a given initial value problem, along with the possibility of handling periodic functions by means of its MOD command.

Finally, DERIVE facilitates to plot the solution and its evaluation at a given point if desired.

Acknowledgments

This work is supported by Spanish Ministry of Education projects MTM2005-01182 and DPI2005-08732-C02-01, with FEDER funds. The authors are grateful to the reviewers for their useful suggestions.

References

[1] R. J. Beerends, H. G. ter Morsche, J. C. van den Berg, and E. M. van de Vrie, *Fourier and Laplace Transforms*, Cambridge University Press, Cambridge, UK, 2003.

[2] J. W. Brown and R. V. Churchill, *Fourier Series and Boundary Value Problems*, McGraw-Hill, New York, NY, USA, 1993.

[3] *DERIVE^{TM}. The Mathematical Assistant for Your Personal Computer*, Texas Instruments, Stafford, Tex, USA, 2000.

[4] M. Legua, J. A. Moraño, and L. M. Sánchez Ruiz, "Generating periodic functions," *WSEAS Transactions on Systems*, vol. 3, no. 1, pp. 37–39, 2004.

[5] S. Wolfram, *The Mathematica® Book*, Wolfram Media/Cambridge University Press, Cambridge, UK, 4th edition, 1999.

[6] *MATLAB®. The Language of Technical Computing*, The MathWorks, Natick, Mass, USA, 2002.

[7] M. Legua, J. A. Moraño, and L. M. Sánchez Ruiz, "Sine and cosine series representations," *WSEAS Transactions on Mathematics*, vol. 3, no. 3, pp. 543–548, 2004.

[8] L. M. Sánchez Ruiz, M. Legua, and J. A. Moraño, *Matemáticas con DERIVE*, Universidad Politécnica de Valencia, Valencia, Spain, 2001.

[9] C. H. Edwards Jr. and D. E. Penney, *Ecuaciones Diferenciales Elementales*, Prentice Hall, Upper Saddle River, NJ, USA, 1993.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 475878, 11 pages
doi:10.1155/2008/475878

Research Article

Detection of Variations of Local Irregularity of Traffic under DDOS Flood Attack

Ming Li[1] and Wei Zhao[2]

[1] School of Information Science and Technology, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, China
[2] Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180-3590, USA

Correspondence should be addressed to Ming Li, ming_lihk@yahoo.com

Received 24 March 2008; Accepted 1 April 2008

Recommended by Cristian Toma

The aim of distributed denial-of-service (DDOS) flood attacks is to overwhelm the attacked site or to make its service performance deterioration considerably by sending flood packets to the target from the machines distributed all over the world. This is a kind of local behavior of traffic at the protected site because the attacked site can be recovered to its normal service state sooner or later even though it is in reality overwhelmed during attack. From a view of mathematics, it can be taken as a kind of short-range phenomenon in computer networks. In this paper, we use the Hurst parameter (H) to measure the local irregularity or self-similarity of traffic under DDOS flood attack provided that fractional Gaussian noise (fGn) is used as the traffic model. As flood attack packets of DDOS make the H value of arrival traffic vary significantly away from that of traffic normally arriving at the protected site, we discuss a method to statistically detect signs of DDOS flood attacks with predetermined detection probability and false alarm probability.

1. Introduction

IP Networks are subject to electronic attacks [1]. An intrusion detection system (IDS) collects information from a variety of systems and network sources to analyze the information of attack signs. A network-based IDS monitors the traffic on its network as a data source [2]. For distributed denial-of-service (DDOS) flood attack, an intruder bombs attack packets upon a site (victim) with a huge amount of traffic the sources of which are distributed over the world [3]. Hence the pattern of traffic under DDOS flood attack may suddenly differ significantly from the normal pattern of the arrival traffic. From the perspective of dynamical aspects for limited time interval in physics [4], one may regard this sudden change as a specific "pulse." Though DDOS flood attack may not be a sole factor to make traffic pattern vary significantly, we assume that secure officers can distinguish significant variation of monitored traffic pattern caused by other known factors (e.g., normally heavy traffic) from DDOS flood attack. Without confusions causing, the term abnormal traffic used in this paper specifically implies a traffic series that has significant variation of traffic pattern caused by DDOS flood attack.

In this research, we ponder two fundamental issues in detection. One is feature extraction of monitored traffic time series. The other is detection scheme that can be used to assure predetermined detection probability (P_d) and false alarm probability (P_f). The first issue will be discussed in Section 2 from a view of feature extraction of traffic based on self-similarity of traffic. The second will be dissertated in Section 3 based on statistical detection. Section 4 will explain the performance analysis of the present detection system. A case study is demonstrated in Section 5. Discussions are given in Section 6, which is followed by conclusions.

2. Feature extraction of traffic

2.1. Self-similar traffic

Computer scientists in the last decade discovered that traffic is a type of fractal time series. It has the properties of self-similarity, long memory, and multiscales (see e.g., [5]). A commonly used model in traffic engineering is fractional Gaussian noise (fGn) (see e.g., [6–8]).

Let $B(t)$, $t \in (0,\infty)$ be Wiener Brownian motion. Let $B_H(t)$ be fractional Brownian motion with the Hurst parameter $H \in (0,1)$ [9]. Let $\Gamma(\cdot)$ be Gamma function. Then by using fractional calculus, $B_H(t)$ is expressed by

$$B_H(t) - B_H(0) = \frac{1}{\Gamma(H+1/2)}\left\{\int_{-\infty}^{0}\left[(t-u)^{H-0.5} - (-u)^{H-0.5}\right]\mathrm{dB}(u) + \int_{0}^{t}(t-u)^{H-0.5}\,\mathrm{dB}(u)\right\}. \tag{2.1}$$

Let $G(t)$ be the increment series of $B_H(t)$:

$$G(t) = B_H(t+a) - B_H(t), \tag{2.2}$$

where a is a real number. Then $G(t)$ is fGn [9]. The autocorrelation function (ACF) of fGn in the discrete case is given by

$$\rho(\tau) = \frac{\sigma^2}{2}\left[\left||\tau|+1\right|^{2H} - 2|\tau|^{2H} + \left||\tau|-1\right|^{2H}\right], \tag{2.3a}$$

where $\sigma^2 = \Gamma(2-H)\cos(\pi H)/\pi H(2H-1)$ is the intensity of fGn [10]. The normalized ACF of fGn is given by

$$R(\tau) = \frac{1}{2}\left[\left||\tau|+1\right|^{2H} - 2|\tau|^{2H} + \left||\tau|-1\right|^{2H}\right]. \tag{2.3b}$$

The relationship between the fractal dimension of fGn and H is given by

$$D = 2 - H. \tag{2.4}$$

Approximating the right side of (2.3b) with the second-order differential of $0.5(\tau)^{2H}$, see [9, H15, page 350], for $\tau \geq 0$, yields

$$0.5\left[(\tau+1)^{2H} - 2\tau^{2H} + (\tau-1)^{2H}\right] \approx H(2H-1)\tau^{2H-2}. \tag{2.5}$$

Let y and R be a traffic series and its ACF, respectively. Then according to (2.5),

$$R(\tau) \sim c\tau^{2H-2}, \quad H \in (0.5, 1), \tag{2.6}$$

where ~ implies the asymptotical equivalence under the limit $\tau \to \infty$ and $c > 0$ is a constant [11].

The ACF (2.5) is nonsummable for $H > 0.5$, implying long-range dependence (LRD). Hence H is a measure of LRD of traffic. It is kindly noted that LRD of traffic does not mean that DDOS attacking is a long-range phenomenon. On the contrary, DDOS attacking and its detection are short-range phenomena since both sides, namely, an attacker and its opponent, are engaged with each other during a short period of time. Such a battle makes local irregularity of traffic vary dramatically [12].

Without losing generality, we consider traffic series y in the discrete case. By dividing y into nonoverlapping blocks of size L and averaging over each block, we obtain another series given by

$$y(i)^{(L)} = \frac{1}{L}\sum_{j=iL}^{(i+1)L} y(j). \tag{2.7}$$

According to the analysis in [5, 9, 11], in the fGn sense, one has

$$\operatorname{Var}\left(y^{(L)}\right) = L^{2H-2}\operatorname{Var}(y), \tag{2.8}$$

where Var implies the variance operator. Thus the self-similarity is measured by H.

A series encountered in engineering is usually of finite length. Let y be a series of P length. Divide it into N nonoverlapping sections. Each section is divided into M nonoverlapping segments. Divide each segment into K nonoverlapping blocks. Each block is of L length. Let $y(i)_m^{(L)}(n)$ be the series with aggregated level L in the mth segment of the nth section ($m = 0, 1, \ldots, M-1;\ n = 0, 1, \ldots, N-1$). Let $H_m(n)$ be the H value of $y(i)_m^{(L)}(n)$. Let $r(k; H_m(n))$ be the measured ACF of $y(i)_m^{(L)}(n)$ in the normalized case. The theoretic ACF form corresponding $y(i)_m^{(L)}(n)$ in the fGn sense is given by

$$R(k; H_m(n)) = 0.5\left[\left||k|+1\right|^{2H_m(n)} - 2|k|^{2H_m(n)} + \left||k|-1\right|^{2H_m(n)}\right]. \tag{2.9}$$

The above expression exhibits the multifractal property of traffic as can be seen from [13].

Let

$$J(H_m(n)) = \sum_k \left[R(k; H_m(n)) - r(k; H_m(n))\right]^2 \tag{2.10}$$

be the cost function. Then one has

$$H_m(n) = \arg\min J\left[H_m(n)\right]. \tag{2.11}$$

Averaging $H_m(n)$ in terms of index m yields

$$H(n) = \frac{1}{M}\sum_{m=0}^{M-1} H_m(n), \tag{2.12}$$

representing the H estimate of the series in the nth section.

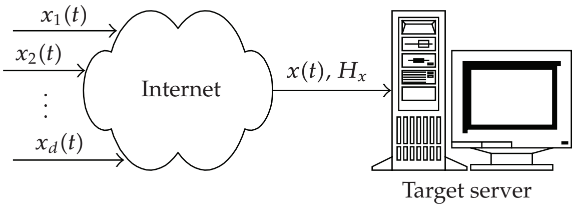


Figure 1: Normal traffic at input of a server.

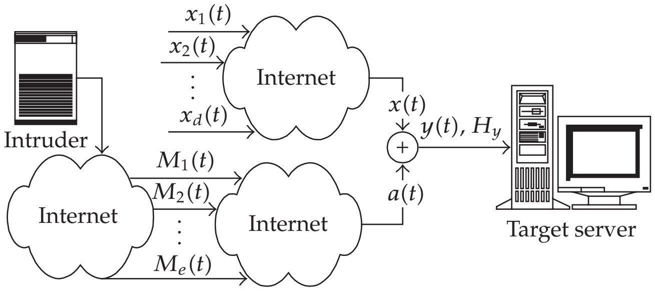


Figure 2: Illustration of abnormal traffic.

Usually, $H(n_1) \neq H(n_2)$ for $n_1 \neq n_2$. However, stationarity of traffic time series implies that $H(n)$ at a specific site is a number falling within a certain confidence interval [5, Paragraph 5, Section 5, page 966]. In practical terms, a normality assumption for $H(n)$ is quite accurate in most cases for $M > 10$ regardless of probability distribution function of H [14]. Thus we take

$$H_x = E[H(n)] \tag{2.13}$$

as a mean estimate of H of x, where E is the mean operator. It can be taken as a template of H of x for the purpose of statistical detection. The appendix gives a case of the H estimation of a real-traffic series to clarify the reasonableness of H in featuring traffic time series.

2.2. Characterizing traffic time series with H

Let x be normal traffic time series. Normally, the site serves x peacefully though x may sometimes be unpleasantly delayed because of the normal traffic jam. The arrival traffic x is contributed by many connections distributed all over the world. Figure 1 shows x contributed by traffic from d connections. From previous discussions, we see that x can be characterized by the Hurst parameter and we denote it as H_x.

Assume that the site is intruded by DDOS flood attacking. Then actual arrival traffic (abnormal traffic) consists of normal traffic x and attack traffic a, see Figure 2, where a is contributed by e connections. We use H_y as a feature of y.

3. Detection method and system structure

To explain our detection principle, we introduce three terms. Correctly recognizing an abnormal sign is termed *detection*; failing to recognize it, *miss*; mistakenly recognizing a normal as abnormal is *a false alarm*.

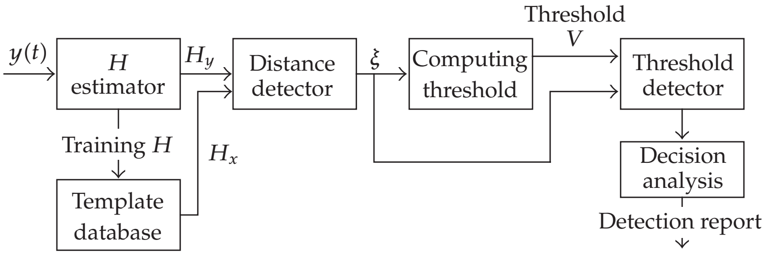


Figure 3: System diagram.

Let $\xi = \|H_x - H_y\|$. Then ξ represents the deviation of H of monitored traffic time series. Let $V > 0$ be the threshold. Then the detection hypotheses are as follows. $\xi > V$, implies detection, while $\varsigma = \|H_x - H_{xl}\| > V$ represents false alarm, where H_{xl} stands for H which is not used as the template but obtained when there is no attacking. Clearly, ξ and ζ are random variables. Mathematically, there are many distance measures available [15–17], but the following works well:

$$\xi = E\left[\sum_k \left|\frac{H_y}{H_x} - \log\frac{H_y}{H_x} - 1\right|\right]. \tag{3.1}$$

According to the previous discussions, we give the system diagram in Figure 3. The measured arrival traffic first passes through an H estimator. The result of H estimator goes to template database to produce the template H_x. In addition, it outputs an online estimate of H_y. H_x and H_y are compared in the distance detector. The comparison result ξ is fed into threshold detector to compare with a given threshold V. In the stage of decision analysis, the output of the threshold detector is analyzed and its output gives a sign of detection according to preset detection probability and false alarm probability.

4. Performance analysis

With the partition explained in Section 2, we see that there is a value of ξ representing the deviation of H of y in each segment. Therefore, in each section, ξ is a random sequence of M length. Denote $\bar{\xi}$ as the expectation of ξ in each section. Then $\bar{\xi}$ is a random sequence of N length. In the case of $N \geq 10$, $\bar{\xi}$ well obeys Gaussian distribution [14]. For the simplicity, we still denote $\bar{\xi}$ as ξ.

4.1. Detection probability

Let μ_ξ and σ_ξ^2 be the expectation and the variance of ξ, respectively. Then

$$\xi \sim N(\mu_\xi, \sigma_\xi^2) = \frac{1}{\sqrt{2\pi}\sigma_\xi} e^{-(\xi-\mu_\xi)^2/2\sigma_\xi^2}. \tag{4.1}$$

Let

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt. \tag{4.2}$$

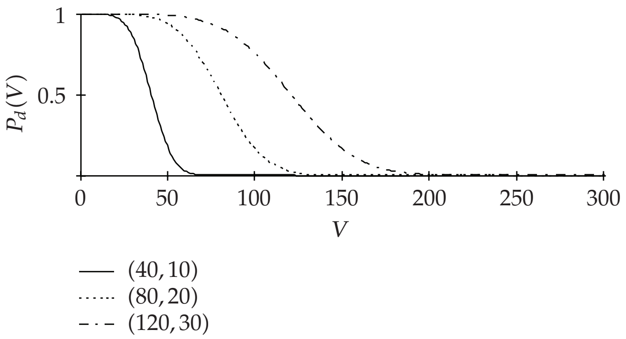


Figure 4: Detection probability.

Then detection probability is given by

$$P_d = P\{V < \xi < \infty\} = \int_{(V-\mu_\xi)/\sigma_\xi}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = 1 - \Phi\left[\frac{V-\mu_\xi}{\sigma_\xi}\right]. \tag{4.3}$$

4.2. False alarm probability

Let μ_ζ and σ_ζ^2 be the mean and the variance of ζ. Then false alarm probability is given by

$$P_f = P\{V < \zeta < \infty\} = \int_{(V-\mu_\zeta)/\sigma_\zeta}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = 1 - \Phi\left[\frac{V-\mu_\zeta}{\sigma_\zeta}\right]. \tag{4.4}$$

4.3. Miss probability

Let P_m be miss probability. Then

$$P_m = P\{-\infty < \xi < V\} = \int_{-\infty}^{(V-\mu_\xi)/\sigma_\xi} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = \Phi\left[\frac{V-\mu_\xi}{\sigma_\xi}\right]. \tag{4.5}$$

Generally, $\mu_\zeta = 0$. Besides, the numeric computation in data processing can be arranged such that $\sigma_\zeta = \sigma_\xi = \sigma$. In this case, three probabilities are given by

$$\begin{aligned} P_d &= \int_{(V-\mu_\xi)/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = 1 - \Phi\left[\frac{V-\mu_\xi}{\sigma}\right], \\ P_f &= \int_{V/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = 1 - \Phi\left(\frac{V}{\sigma}\right), \\ P_m &= \int_{-\infty}^{(V-\mu_\xi)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = 1 - \Phi\left[\frac{V-\mu_\xi}{\sigma}\right]. \end{aligned} \tag{4.6}$$

Figures 4–6 show the curves of three distributions, respectively. As $P_d + P_m = 1$, high P_d implies low P_m and vice versa.

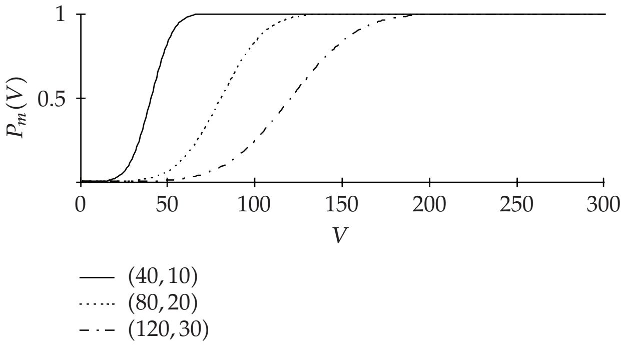


Figure 5: Miss probability.

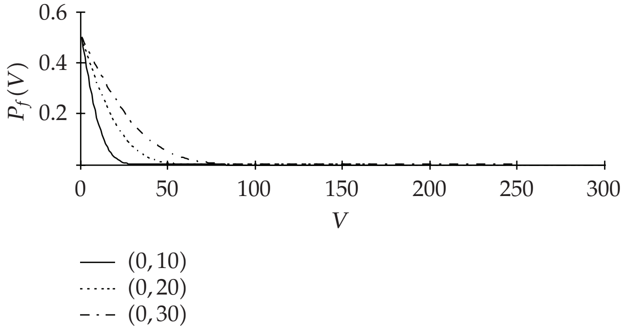


Figure 6: False alarm probability.

4.4. Threshold and detection region

As can be seen from the previous discussions, the selection of a threshold value is crucial to our system. In fact, given a false alarm probability f, we want to find the threshold V_f such that $P(V_f) \leq f$. Clearly,

$$V_f \geq -\sigma\Phi^{-1}(f). \tag{4.7}$$

If $f = 0$ and when the selected precision is 4, we obtain

$$V_f \geq 4\sigma. \tag{4.8}$$

Given a detection probability d, we want to find the threshold V_d such that $P_d(V_d) \geq d$. Clearly,

$$V_d \leq \mu_\xi - \sigma\Phi^{-1}(d), \quad \text{if } \mu_\xi - \sigma\Phi^{-1}(d) > 0. \tag{4.9}$$

In the case of $d = 1$,

$$V_d \leq \mu_\xi - 4\sigma, \quad \text{if } \mu_\xi - 4\sigma > 0. \tag{4.10}$$

Therefore, when $-\sigma\Phi^{-1}(f) < \mu_\xi - \sigma\Phi^{-1}(d)$ and $V \in [-\sigma\Phi^{-1}(f), \mu_\xi - \sigma\Phi^{-1}(d)]$, $P_d \geq d$ and $P_f \leq f$ are assured. That is,

$$\begin{matrix} P_d \geq d, \\ P_f \leq f, \end{matrix} \quad \text{if } V \in \left[-\sigma\Phi^{-1}(f), \mu_\xi - \sigma\Phi^{-1}(d)\right],\ \mu_\xi - \sigma\Phi^{-1}(d) > 0. \tag{4.11}$$

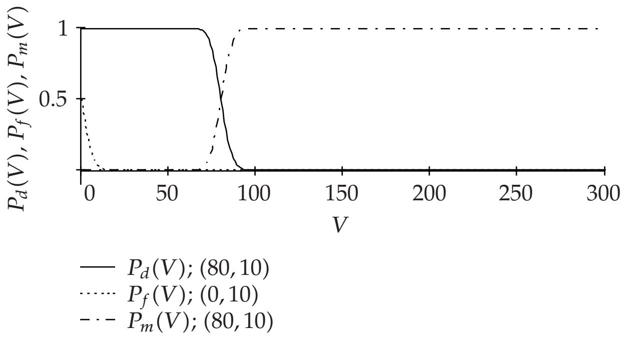


Figure 7: Intersection of three probability distributions: detection region.

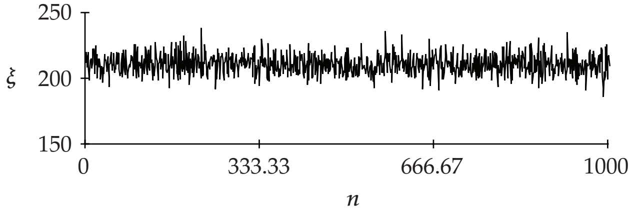


Figure 8: Random variable ξ.

In the case of $d = 1$ and $f = 0$,

$$\begin{aligned} P_d &= 1, \\ P_f &= 0, \end{aligned} \quad \text{if } V \in [4\sigma, \mu_\xi - 4\sigma],\ \mu_\xi - 4\sigma > 0. \tag{4.12}$$

The constraint of (4.12) is given by $\mu_\xi > 8\sigma$.

Obviously, the detection region is the intersection of three probability functions. Under the condition of $\mu_\xi = 80$ and $\sigma = 10$, the detection region is shown in Figure 7.

5. A case study

Suppose the template $H_0 = 0.7671$ as described in the appendix. Assume that the confidence level is 99.9999%. Thus we suppose y's $H \in (0.5000, 0.7669)$ or $(0.7673, 0.9900)$ during the transition process of intrusion. In this case study, 1000 points of Hs in $(0.5000, 0.7669)$ or (0.7673,0.9900) are randomly selected to simulate the abnormal traffic deviating from the normal one. The error sequence is indicated in Figure 8. By the numeric computation, we obtain $\mu_\xi = 210.3011$ and $\sigma = 7.7490$. Therefore, we obtain the probability distributions for detection, false alarm and miss as shown in Figure 9. Under the conditions of $P_d = 1$ and $P_f = 0$, we obtain $V_{\min} = 30.9951$ and $V_{\max} = 179.3052$. Hence when we select $V \in [30.9951, 179.3052]$, we have 99.9999% confidence to say that $P_d = 1$ and $P_f = 0$ are assured, which can be easily observed from Figure 9.

6. Discussions

Since Yahoo servers were successfully attacked in 2001, the issue of detecting DDOS flood attacking has been paid much attention to. Various methods and systems have been

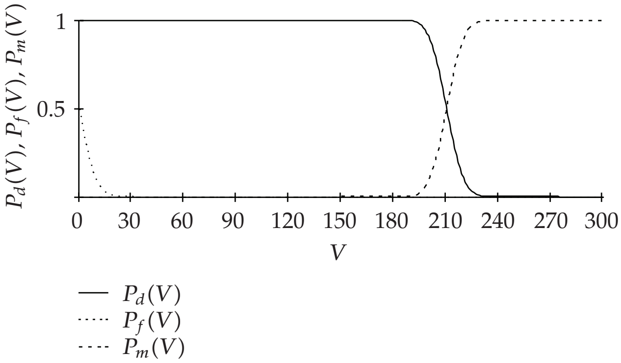


Figure 9: Case study: detection region.

proposed, see, for example, [18–25]. As known, traffic under DDOS flood attack must be significantly different from that of normal one [25]. Otherwise, DDOS flood attack would have no effect. From this point of view, the value of H of traffic under DDOS flood attacks is considerably different from that of normal one, see [12] for details.

For a stationary random time series of finite length, ACF and power spectrum density (PSD) function are commonly used in engineering for feature extraction in statistical classifications [16, 17]. However, the PSD of traffic does not exist in the domain of ordinary functions since it has long memory [8]. To avoid such a difficulty in mathematics, consequently, ACF of traffic is considered for feature extraction in our early work [25]. This paper focuses on detection of local variations of traffic based on the self-similarity of traffic. Thus it suggests a new method that substantially develops the work of [25], from the point of view of traffic pattern matching, because feature extraction of traffic time series by using a single parameter H makes pattern matching more efficient.

7. Conclusions

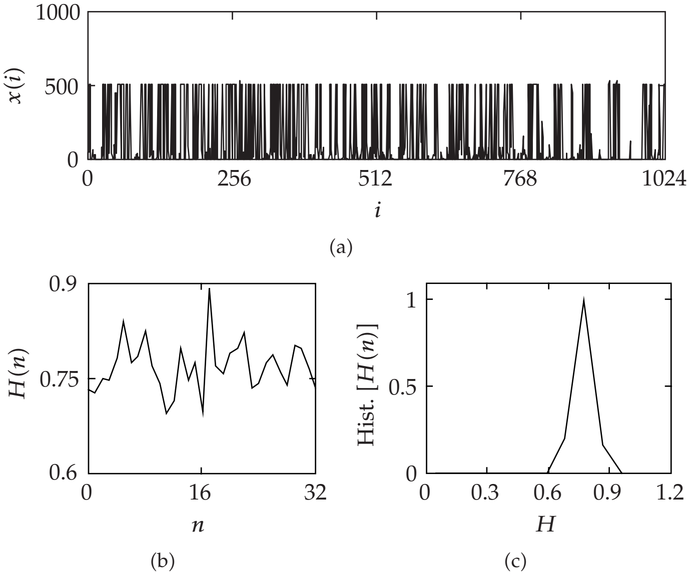
We have discussed the characterization of the local irregularity of traffic by $H(n)$. We have explained a principle of statistical detection to capture signs of DDOS flood attacking with predetermined detection probability and false alarm probability based on the variation of the local irregularity of traffic.

Appendix

Demonstration of H estimation of a real-traffic series

This appendix gives a demonstration with a real-traffic series, named LBL-PKT-4 [26, 27]. Denote $x(i)$ as the series of LBL-PKT-4, indicating the number of bytes in the ith packet. The length of that series is 1.3 million. The first 1024 points of that series is plotted in Figure 10(a). Divide $x(i)$ into 32 nonoverlapping sections. Computing H in each section yields $H(n)$ ($n = 0, 1, \ldots, 31$) as shown in Figure 10(b). Its histogram is indicated in Figure 10(c).

According to (2.13), we have $H_x = 0.7671$. The confidence interval with 95% confidence level is [0.7670,0.7672]. Hence we have 95% confidence to say that the H estimate in each section of that series takes $H_x = 0.7671$ as its approximation with fluctuation not greater than 1×10^{-4}. Further, it is easy to obtain that the confidence interval with 99.9999%



(a)

(b)

(c)

Figure 10: Verification of statistical invariable H. (a) A real-traffic time series; (b) estimate $H(n)$; (c) histogram of $H(n)$.

confidence level is [0.7669, 0.7673]. Hence we have 99.9999% confidence to say that the H estimate in each section of that series takes $H_x = 0.7671$ as its approximation with fluctuation not greater than 2×10^{-4}.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under the project Grant no. 60573125. Wei Zhao's work was also partially supported by the NSF (USA) under Contracts no. 0808419, 0324988, 0721571, and 0329181. Any opinions, findings, conclusions, and/or recommendations in this paper, either expressed or implied, are those of the authors and do not necessarily reflect the views of the agencies listed above.

References

[1] G. Coulouris, J. Dollimore, and T. Kindberg, *Distributed Systems: Concepts and Design*, Addison-Wesley, Reading, Mass, USA, 3rd edition, 2001.

[2] E. G. Amoroso, *Intrusion Detection: An Introduction to Internet Surveillance, Correlation, Traps, Trace Back, and Response*, Intrusion.Net Book, Sparta, NJ, USA, 1999.

[3] L. Garber, "Denial-of-service attacks rip the Internet," *Computer*, vol. 33, no. 4, pp. 12–17, 2000.

[4] G. Toma, "Practical test functions generated by computer algorithms," in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA '05)*, vol. 3482 of *Lecture Notes in Computer Science*Lecture Notes in Computer Science, pp. 576–584, Singapore, May 2005.

[5] W. Willinger and V. Paxson, "Where mathematics meets the Internet," *Notices of the American Mathematical Society*, vol. 45, no. 8, pp. 961–970, 1998.

[6] M. Li, W. Zhao, W. Jia, D. Long, and C.-H. Chi, "Modeling autocorrelation functions of self-similar teletraffic in communication networks based on optimal approximation in Hilbert space," *Applied Mathematical Modelling*, vol. 27, no. 3, pp. 155–168, 2003.

[7] B. Tsybakov and N. D. Georganas, "Self-similar processes in communications networks," *IEEE Transactions on Information Theory*, vol. 44, no. 5, pp. 1713–1725, 1998.

[8] A. Adas, "Traffic models in broadband networks," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 82–89, 1997.

[9] B. B. Mandelbrot, *Gaussian Self-Affinity and Fractals*, Springer, New York, NY, USA, 2002.

[10] M. Li and S. C. Lim, "A rigorous derivation of power spectrum of fractional Gaussian noise," *Fluctuation and Noise Letters*, vol. 6, no. 4, pp. C33–C36, 2006.

[11] J. Beran, *Statistics for Long-Memory Processes*, vol. 61 of *Monographs on Statistics and Applied Probability* Monographs on Statistics and Applied Probability, Chapman and Hall, New York, NY, USA, 1994.

[12] M. Li, "Change trend of averaged Hurst parameter of traffic under DDOS flood attacks," *Computers & Security*, vol. 25, no. 3, pp. 213–220, 2006.

[13] M. Li and S. C. Lim, "Modeling network traffic using generalized Cauchy process," *Physica A*, vol. 387, no. 11, pp. 2584–2594, 2008.

[14] J. S. Bendat and A. G. Piersol, *Random Data. Analysis and Measurement Procedures*, John Wiley & Sons, New York, NY, USA, 3rd edition, 2000.

[15] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal Processing*, vol. 18, no. 4, pp. 349–369, 1989.

[16] K. S. Fu, Ed., *Digital Pattern Recognition*, Springer, Berlin, Germany, 2nd edition, 1980.

[17] A. R. Webb, *Statistical Pattern Recognition*, Edward Arnold, London, UK, 1999.

[18] M. Li and W. Zhao, "A statistical model for detecting abnormality in static-priority scheduling networks with differentiated services," in *Proceedings of the International Conference on Computational Intelligence and Security (CIS '05)*, vol. 3802 of *Lecture Notes in Computer Science*Lecture Notes in Computer Science, pp. 267–272, Springer, Xi'an, China, December 2005.

[19] V. Paxson, "Bro: a system for detecting network intruders in real time," in *Proceedings of the 7th USENIX Security Symposium*, San Antonio, Tex, USA, January 1998.

[20] W. Yu, D. Xuan, and W. Zhao, "Middleware-based approach for preventing distributed deny of service attacks," in *Proceedings of IEEE Military Communications Conference (MILCOM '02)*, vol. 2, pp. 1124–1129, Anaheim, Calif, USA, October 2002.

[21] P. Innella and O. McMillan, "An introduction to intrusion detection systems, tetrad digital integrity, LLC," December 2001, http://www.securityfocus.com/infocus/1520/.

[22] http://en.wikipedia.org/wiki/Denial-of-service_attack/.

[23] http://www.sans.org/dosstep/index.php/.

[24] R. Bettati, W. Zhao, and D. Teodor, "Real-time intrusion detection and suppression in ATM networks," in *Proceedings of the 1st USENIX Workshop on Intrusion Detection and Network Monitoring*, Santa Clara, Calif, USA, April 1999.

[25] M. Li, "An approach to reliably identifying signs of DDOS flood attacks based on LRD traffic pattern recognition," *Computers & Security*, vol. 23, no. 7, pp. 549–558, 2004.

[26] http://www.acm.org/sigcomm/ITA/.

[27] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling," *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 226–244, 1995.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 510406, 15 pages
doi:10.1155/2008/510406

Research Article

Tool Wear Detection Based on Duffing-Holmes Oscillator

Wanqing Song,[1] Shen Deng,[2] Jianguo Yang,[1] and Qiang Cheng[2]

[1] *College of Mechanical Engineering, Donghua University, Shanghai 201620, China*
[2] *Shanghai University of Science, 333#, Longteng Road, Songjiang district, Shanghai 201620, China*

Correspondence should be addressed to Wanqing Song, swqls@126.com

Received 28 May 2008; Accepted 27 July 2008

Recommended by Carlo Cattani

The cutting sound in the audible range includes plenty of tool wear information. The sound is sampled by the acoustic emission (AE) sensor as a short-time sequence, then worn wear can be detected by the Duffing-Holmes oscillator. A novel engineering method is proposed for determining the chaotic threshold of the Duffing-Holmes oscillator. First, a rough threshold value is calculated by local Lyapunov exponents with a step size 0.1. Second, the exact threshold value is calculated by the Duffing-Holmes system in terms of the law of the golden section. The advantage of the method is low computation cost. The feasibility for tool condition detection is demonstrated by the 27 kinds of cutting conditions with sharp tool and worn tool in turning experiments. The 54 group data sampled as noisy are embedded into the Duffing-Holmes oscillator, respectively. Finally, one chaotic threshold is determined conveniently which can distinguish between worn tool or sharp tool.

1. Introduction

Tool wear is a complex phenomenon occurring in metal cutting processes. A worn tool adversely affects the surface finish of the work piece and therefore there is a need to detect tool wear which alerts the operator to the tool wear state, thus, avoiding undesirable product quality. However, accurately determining cutting conditions remains difficult.

Acoustic emission based on tool condition monitoring has been available for approximately 17 years, most of them use analog root mean square of the signal to monitor tool wear or detect breakages. Damodarasamy and Raman [1] combined the radial force, feed force, and AE to model the tool flank wear in a turning operation. Wanqing et al. [2] used a wavelet transform and fractal algorithm to capture the features of the AE signals. Yao et al. [3] used a fuzzy neural network to describe the relation between the monitoring features, which are derived from wavelet-based AE signals, and the tool wear condition. The data processing

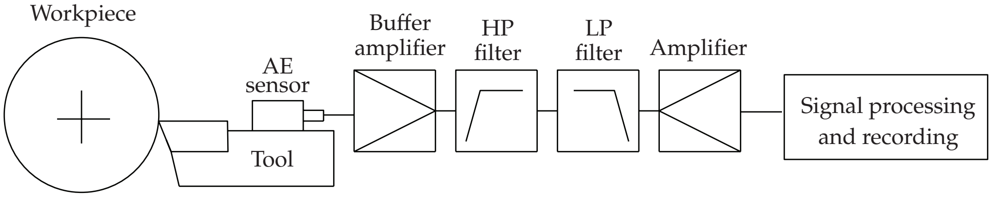


Figure 1: AE measurement in metal cutting.

methods have shown acoustic emission signal power to increase with tool wear owing to increased friction effect [4].

Nearly years, chaotic oscillator is used widely to detect weak period signal [5–8]. The weak signal detection is a central problem in the general field of signal processing and the use of chaos theory in weak signal detection is also a topic of interest in chaos control. At present, however, this research is mainly theory and simulation, engineering practice is a few examples. The phase transforms of Duffing-Holmes oscillator are sensitive to periodic signal and periodic interference signals which have larger angular frequency difference from the referential signal, but immune to the random noisy [5, 9]. Since tool wear is a gradual processing during the turning conditions, the cutting sound is composed of periodic signals and a large amount of periodic interference signals and the random noise. Of course, the frequency and amplitude of these signals also are changing gradually along with tool wear except of the random noise. Therefore, the tool wear processing belongs to detect weak periodic signals in strong noisy and very appropriately by Duffing-Holmes oscillator.

Machining tests were carried out on HL-32 NC turning center. This lathe does not have a tailstock. Tungsten carbide finishing tool was used to turn free machining mild steel. The work material was chosen for ease of machining, allowing for generation of surfaces of varying quality without the use of cutting fluids. The experiment equipments are shown in Figure 1. The piezoelectric AE sensor (CAE-150) was mounted on the tool holder. A light coating of petroleum jelly was applied under the sensor to ensure good acoustic emission coupling. Because of high impedance of the sensor, it must be directly connected to a buffer amplifier. Low-frequency noise components, which are inevitably present in AE signal, cannot represent the tool's condition and hence useless. Therefore, those components should be eliminated (highpass filtered) at the earliest possible stage of signal processing to enable usage of full amplitude range of the equipment. The filtered signals were sampled at 4 MHz using a digital storage oscillograph to a PC, see Figure 1. All test data were processed and analyzed by using the Matlab software.

In the experiment, according to the cutting conditions which are presented in Table 1, a sharp tool and a worn tool was used, respectively.

The data sampled by AE, 54 group data, are merged into Duffing-Holmes equation as an exterior perturbation of the chaotic system, respectively. Then, with tool wear, the gradual change sound signal under the background of strong noise can be detected by identifying the phase space trajectory. In terms of the results from theoretical calculation, it is proved that there is a huge difference in the phase space trajectories between the chaotic state and the periodic state, and this difference can be used as the evidence in the chaotic system for the detection of tool wear signal based on Duffing oscillator. Meanwhile, Lyapunov exponents are adopted as threshold value evaluated roughly for chaotic critical state, the law of golden section to determine the threshold is proposed and the threshold in chaotic critical state is

Table 1: Experimental cutting conditions.

No.	Speed	Depth of cut	Feed	No.	Speed	Depth of cut	Feed
	r/min	mm	mm/r		r/min	mm	mm/r
1	1500	1	0.1	15	1000	0.2	0.05
2	1500	0.5	0.1	16	800	1	0.05
3	1500	0.2	0.1	17	800	0.5	0.05
4	1000	1	0.1	18	800	0.2	0.05
5	1000	0.5	0.1	19	1500	1	0.02
6	1000	0.2	0.1	20	1500	0.5	0.02
7	800	1	0.1	21	1500	0.2	0.02
8	800	0.5	0.1	22	1000	1	0.02
9	800	0.2	0.1	23	1000	0.5	0.02
10	1500	1	0.05	24	1000	0.2	0.02
11	1500	0.5	0.05	25	800	1	0.02
12	1500	0.2	0.05	26	800	0.5	0.02
13	1000	1	0.05	27	800	0.2	0.02
14	1000	0.5	0.05				

evaluated more accurately. Melniko's function also can be used to calculate the threshold for chaos, but Melniko's function only determines the threshold from order to chaos, but Lyapunov exponents can determine the threshold from chaos to order [7, 10, 11]. We describe a means for tool wear whether or not a system is chaotic. When the tool is sharp, the Duffing-Holmes oscillator is chaos in state space trajectory, when the tool is wear, the Duffing-Holmes oscillator takes on periodic trajectory from chaos to order in state space.

2. Principle detecting weak signal based on Duffing-Holmes oscillator

The Duffing-Holmes is the second differential equation containing the item of the power five, which can be motivated by exterior stimulations to engender oscillation movement and then generate chaotic trajectory or periodic trajectory; its dynamic equation is as follows:

$$x''(t) + 0.5x'(t) - x^3(t) + x^5(t) = r\sin(t) + (\text{input}), \tag{2.1}$$

where 0.5 denotes the ratio of damping, $r\sin(t)$ is the forced periodic terms, which is the reference signal and as an internal signal, $-x^3(t) + x^5(t)$ term is the nonlinear recovery force in system 1, the kinematical state of the system mainly depends on this recovery force term $r\sin(t)$. *Input* terms are the signal measured which is imported to the dynamic system as the supplement of special parameters of chaotic oscillator; we can adjust the amplitude r of the reference signal to the special value as in the chaotic critical state. The value is called threshold value in the chaotic system 1. If a weak periodic signal is merged into system 1, so long as the threshold is adjusted appropriately, the behavior of the Duffing-Holmes will be changed dramatically from chaotic states to periodic states. For example, let *input* terms be $f(t) = 0.2\ \sin(t)$, then the Duffing-Holmes equation is

$$x''(t) + 0.5x'(t) - x^3(t) + x^5(t) = r\sin(t) + f(t), \tag{2.2}$$

when $f(t)$ contains a weak white noisy, that is, $f(t) = 0.2\,\sin(t) + 0.01$ rand, rand is a random white noisy $(0 \sim 1)$, *input* terms $f(t)$ are a low-amplitude periodicsignal with white noisy, then the Duffing-Holmes equation is

$$x''(t) + 0.5x'(t) - x^3(t) + x^5(t) = r\sin(t) + f(t), \tag{2.3}$$

when one weak periodic noisy signal is merged into *input* terms $f(t)$, that is,

$$\begin{aligned} \widehat{f_i} &= f_i, \quad i \neq 16 + 128k, \\ \widehat{f}_{16+128k} &= f_{15+128k} + 0.04, \quad k = 0, 1, 2, 3, \ldots, \end{aligned} \tag{2.4}$$

then

$$x''(t) + 0.5x'(t) - x^3(t) + x^5(t) = r\sin(t) + f(t). \tag{2.5}$$

Let dynamical system 2, 3, and 4 initial point $x'(0) = 0$, $x(0) = 0$, then set threshold $r_d = 0.52544$ as the critical state for the system 2, integrated with Runge-Kutta method of fourth order with a fixed step size $t = 0.01$ second. Total time is 16 seconds. The phase space in systems 2 and 3 takes on periodic state trajectory, but phase space in system 4 is chaotic trajectory.

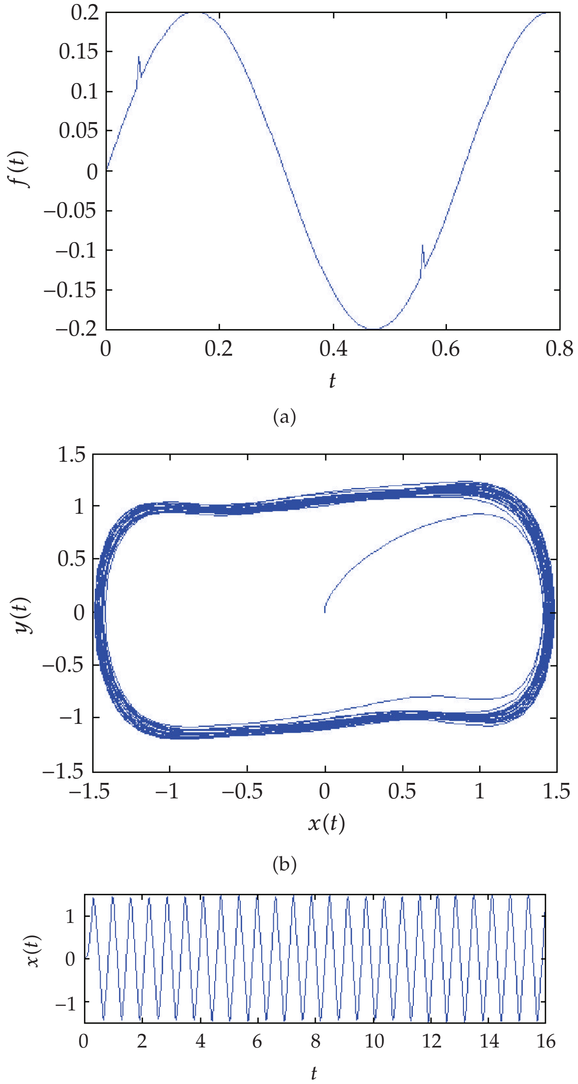
When a strong noise $[0.2\,\sin(t)]$ without white noisy or with white noisy is added to the Duffing-Holmes system, both systems 2 and 3 take on the periodic state. It means the random noisy is not influenced on the state of the dynamic system. Once the strong noisy contains a weak periodic noise signal, the behaviors of system 4 is changed immediately from a large-scale periodic state to a chaotic state. The temporal waveform of $f(t)$, the phase orbit, and the temporal waveforms of systems 3 and 4 are shown in Figures 2 and 3. In other words, the Duffing-Holmes takes on some immunity to random noisy [12] and strong sensitivity to some weak periodic signal. Since 0.01 rand term is too small, it is not obvious in the temporal waveform.

3. Threshold calculated based on Lyapunov exponents

Lyapunov exponents are frequently computed measure for the characteristic of chaotic dynamics [10, 11, 13]. It describes a method for diagnosing whether or not a system is chaotic. To confirm the existence of the weak periodic signal to be detected and the amplitude of the signal, we need to define a proper index for denoting the change in the states of the chaos detection system. The index should be sensitive to a weak periodic signal, but insensitive to the random noise from the viewpoint of statistical characteristics. Thus, the dynamic properties of a certain system are reflected statistically by Lyapunov exponents which are described as follows [14–16].

Dynamic system $x''(t) + 0.5x'(t) - x^3(t) + x^5(t) = r\cos(t)$ is transformed below:

$$\begin{aligned} y(t) &= x'(t), \\ y'(t) &= -0.5\,y(t) + x^3(t) - x^5(t) + r\cos(t). \end{aligned} \tag{3.1}$$



(a)

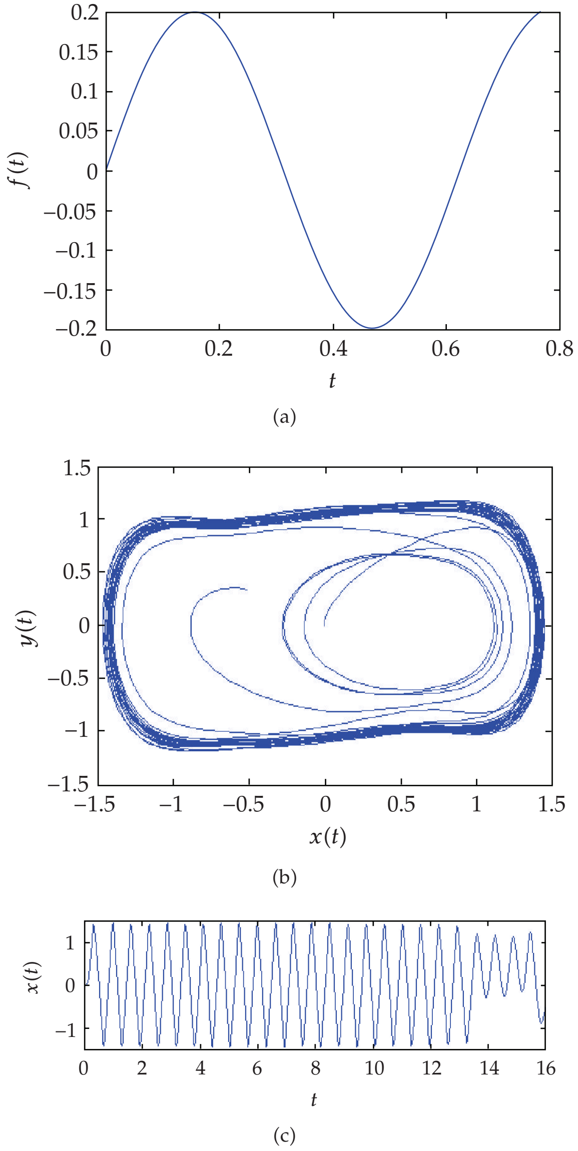
(b)

(c)

Figure 2: Dynamic character with weak periodic noisy signal, $y(t) = x'(t)$.

To a two-dimensional plane $x(t),\ y(t) = x'(t)$, two Lyapunov exponents can be solved in system 5. When the system is in the large-scale periodic state, both of the two Lyapunov exponents are negative. When the system is in the chaotic state, at least one of the two Lyapunov exponents of the system is positive which has behaviors of the chaos. Therefore, the detection system is established on the basis of Lyapunov exponents.

Let initial condition $x(0) = 1,\ x'(0) = 1$, with about typical 30 points in the region $r = [0.5, 1]$ chosen to calculate the Lyapunov exponents (LE), the computation precision of r is two digits after the decimal dot, see Table 2. LE curve are plotted in Figure 4. $r = 0.70$, system 5 takes on the chaotic state, and $r = 0.78$, system 5 takes on the periodic state. They are shown in Figures 5 and 6.



(a)

(b)

(c)

Figure 3: Dynamic character without weak periodic noisy signal, $y(t) = x'(t)$.

Obviously, LE changed from positive to negative correspond to region $r = [0.733, 0.734]$ based on the chaotic system extreme sensitivity to parameters changed. If the threshold r is equal to 0.733, because computation precision of r is only three effective digits after decimal dot, the sensitivity from chaos to periodic is not enough. Above, computation cost spends about 3 hours for typical 30 point sets of r with Matlab. In order to improve sensitivity of system 5, however, if the computation precision of r is risen 4 digits after decimal dot, namely, $r = [0.5200, 0.9800]$, time interval 0.01 second and 1000 steps, the computation cost will spend about 30 hours with Matlab. The more high sensitivity is, the more long computation time is.

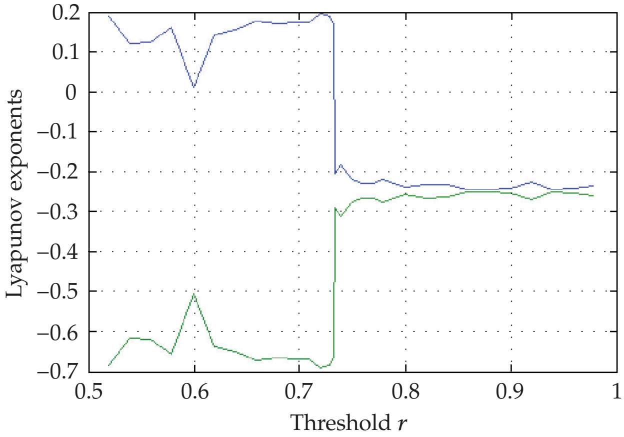


Figure 4: The relational curve of LE and r.

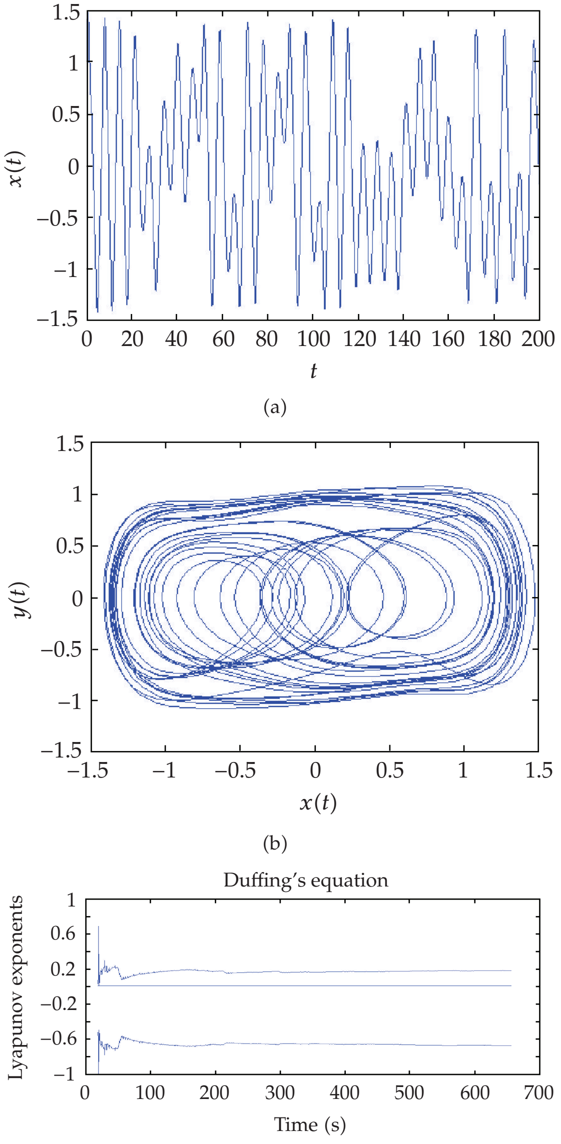
Table 2: Lyapunov exponents in Duffing-Holmes.

No.	r	Max LE	Min LE	No.	r	Max LE	Min LE
1	0.52	0.185	−0.685	16	0.74	−0.18527	−0.31473
2	0.54	0.11914	−0.61914	17	0.75	−0.22179	−0.27821
3	0.56	0.12314	−0.62314	18	0.76	−0.23059	−0.26941
4	0.58	0.15765	−0.65765	19	0.77	−0.23088	−0.26912
5	0.6	0.00782	−0.50782	20	0.78	−0.22217	−0.27783
6	0.62	0.13867	−0.63867	21	0.8	−0.24112	−0.25888
7	0.64	0.15349	−0.65349	22	0.82	−0.23407	−0.26893
8	0.66	0.17321	−0.67321	23	0.84	−0.23449	−0.26551
9	0.68	0.16921	−0.66921	24	0.86	−0.24647	−0.25353
10	0.7	0.17056	−0.67056	25	0.88	−0.24773	−0.25227
11	0.71	0.17226	−0.67226	26	0.9	−0.24394	−0.25606
12	0.72	0.19317	−0.69317	27	0.92	−0.22882	−0.27118
13	0.73	0.185	−0.685	28	0.94	−0.24693	−0.25307
14	0.733	0.164	−0.664	29	0.96	−0.24271	−0.25729
15	0.734	−0.20655	−0.29345	30	0.98	−0.23848	−0.26152

4. Threshold computation combined the law of golden section with Lyapunov exponents

First, rough region of the system threshold r is estimated by Lyapunov exponents with computation precision to be one digit after decimal dot, the calculating process only spends about 40 minutes in the region $r = [0,1]$ with step size 0.1. Whatever any kinds of weak external signal merged, the region of $r = [0.7, 0.8]$ is always sensitivity region changed from chaotic state to large periodic state in system 5. Since the law of the golden section can search optimizing solution quickly [12], the threshold value is determined by the golden section accurately in the region $r = [0.7, 0.8]$. The Duffing-Holmes oscillator is below:

$$x''(t) + 0.5x'(t) - x^3(t) + x^5(t) = r\cos(t). \tag{4.1}$$



(a)

(b)

(c)

Figure 5: $r = 0.70$, system character and LE.

Let initial condition $x(0) = 1,\ x'(0) = 1$, the computation precision of threshold value is six digits after the decimal dot in system 6. The method is as follows:

(1) because 0.7 corresponds to chaotic state and 0.8 corresponding periodic state, $r = 0.75$ is the middle value between 0.7 and 0.8.

(2) because $r = 0.75$ corresponds to periodic states, the region of r is $[0.7, 0.75]$. Then, r is accumulated from 0.7 to 0.75 with the step 0.01 up to 0.71 which corresponds to chaotic state and 0.72 which corresponds to periodic state. 0.715 is the middle value between 0.71 and 0.72.

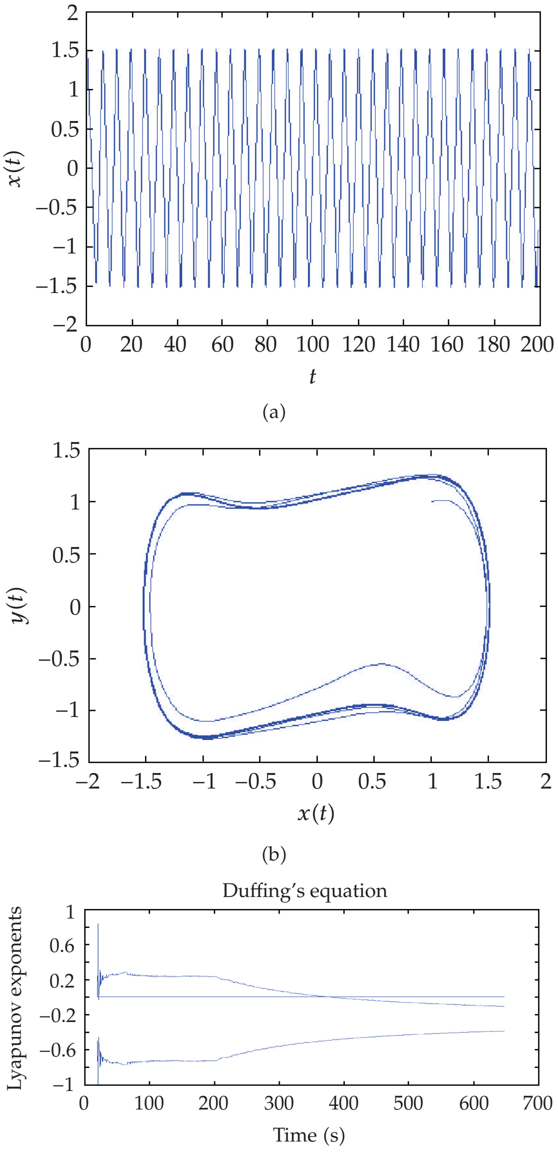
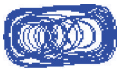
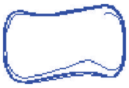
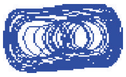
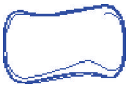
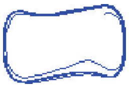
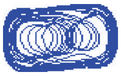
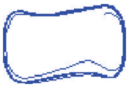
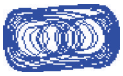
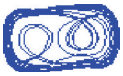
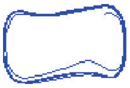
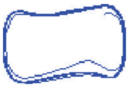
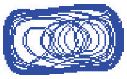
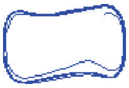
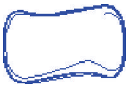
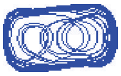
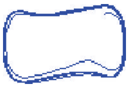
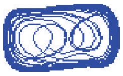


Figure 6: $r = 0.78$, system character and LE.

(3) because $r = 0.715$ corresponds to chaotic state, the region of r is taken $[0.715, 0.72]$. Then, r is accumulated from 0.715 to 0.72 with the step 0.001 up to $r = 0.717$ which corresponds to chaotic state and 0.718 which corresponds to periodic state, 0.7175 is the middle value between 0.717 and 0.718.

(4) because $r = 0.7175$ corresponds to periodic state, the region of r is $[0.717, 0.7175]$. Then, r is accumulated from 0.717 to 0.7175 with the step 0.0001 up to 0.7173 which corresponds to chaotic state and 0.7174 which corresponds to periodic state. 0.71735 is the middle value between 0.7173 and 0.7174.

Table 3: Threshold r based on the golden section in the region $r = [0.7, 0.8]$.

Chaos, r_1	Phase plane	Periodic, r_2	Phase plane	Golden section, r	Phase plane
0.7		0.8		$\frac{r_1 + r_2}{2}$	
0.71		0.72		0.75	
0.717		0.718		0.715	
0.7173		0.7174		0.7175	
0.71732		0.71733		0.71735	
0.717329		0.717330		0.717325	

(5) because $r = 0.71735$ corresponds to periodic state, the region of r is $[0.7173, 0.71735]$. Then, r is accumulated from 0.7173 to 0.71735 with the step 0.00001 up to 0.71732 which corresponds to chaotic state and 0.71733 which corresponds to periodic state. 0.717325 is the middle value between 0.71732 and 0.71733.

(6) because $r = 0.717325$ corresponds to periodic state, the region of r is $[0.717325, 0.71733]$. Then, r is accumulated from 0.717325 to 0.71733 with the step 0.000001 up to 0.717329 which corresponds to chaotic state and 0.717330 which corresponds to periodic state.

(7) final, the threshold value calculated is 0.717329. When a weak periodic signal is merged into system 6, the system takes on the large-scale periodic state. Calculating process is shown in Table 3.

The computation processing only spends about 10 minutes for computation precision to be six digits after the decimal dot. The method has important meaning for engineering practice. 30th steps calculated yield the search optimization threshold value. This is the most amounts of the point sets in the case.

5. Experiment work

The sound signal of sharp tool sampled by AE as an initial condition is merged into the Duffing-Holmes system 6 which is in the chaotic critical state, (its phase plane changes from the chaotic state to the large-scale periodic state), the movement state of the system will

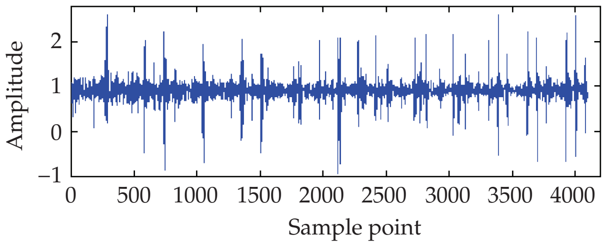


Figure 7: Waveform of sharp tool in first condition.

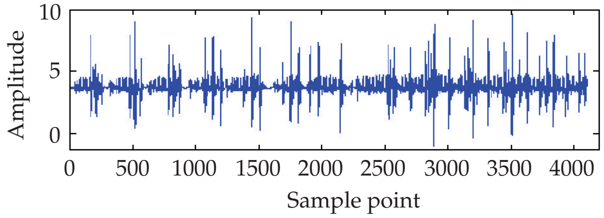


Figure 8: Waveform of wear tool in first condition.

transit immediately from the chaotic state to the large-scale periodic state. The simulation of systems 3 an 4 above has only one input signal, however, for this practice engineering, since the sharp tool and wear tool have 27 groups data, respectively, see systems 7 and 8, the threshold in the both the systems must satisfy to distinguish sharp tool and wear tool in 54 group data. The dynamic system 6 is transformed to systems 7 and 8. When the data of sharp tool are embedded to the chaotic system 7, the phase space is chaotic state; however, when the data of wear tool are embedded to the chaotic system 8, phase space change is the large-scale periodic state. The method based on the change of the dynamic behaviors of a chaotic system (chaotic state, periodic state) has been proposed for recognizing, where there exists a signal to be detected in a system, and greatly immune to the random noise of arbitrary zero average value with unknown probability distribution. The threshold value r should firstly be determined in system 7, which is the critical problem of wear signal chaotic detection. The algorithm to determine the threshold value, using Lyapunov exponents method based on the golden section is detailed as follows:

$$\begin{aligned} x''(t) + 0.5x'(t) - x^3(t) + x^5(t) &= r\cos(t) + \text{sharp\_1}, \\ x''(t) + 0.5x'(t) - x^3(t) + x^5(t) &= r\cos(t) + \text{wear\_1}. \end{aligned} \tag{5.1}$$

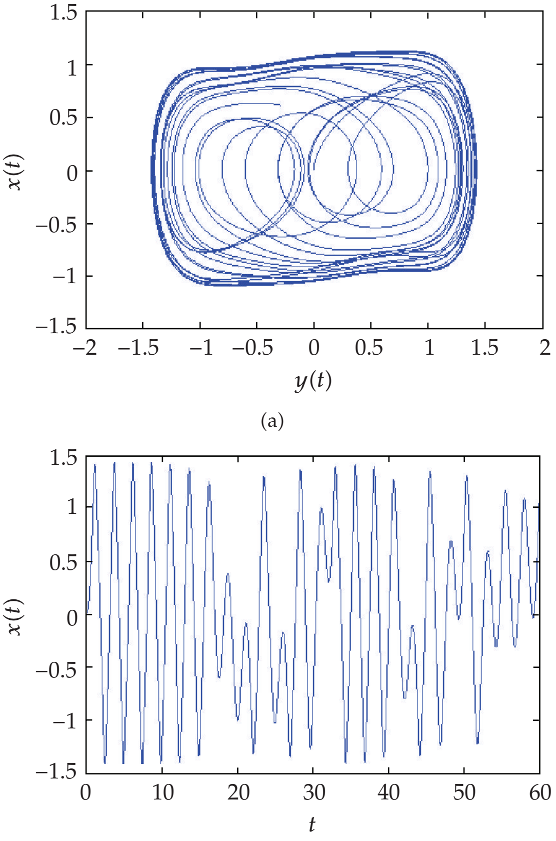
Since the signal amplitude merged is too bigger than interior perturbation force r, the signal sampled is decreased 100 times, thus, signals embedded to Duffing-Holmes are weak perturbation noisy, see systems 9 and 10. The interior perturbation force r is still main signal in the dynamic systems 9 and 10. Sharp_1 signal and wear_1 are shown in Figures 7 and 8 in time domain the initial condition is $[0, 0]$ in systems 9 and 10 frequency sampled is 0.001 second. We set up a chaotic oscillator sensitive to weak periodic signals based on the Duffing-Holmes equation (5.2), and poising the system at its critical state

$$x''(t) + 0.5x'(t) - x^3(t) + x^5(t) = r\cos(t) + 0.01\ \text{sharp\_1}, \tag{5.2}$$

$$x''(t) + 0.5x'(t) - x^3(t) + x^5(t) = r\cos(t) + 0.01\ \text{wear\_1}. \tag{5.3}$$

Table 4: The critical value for 27 group sharp tool data.

No.	r	No.	r	No.	r
1	0.724896	10	0.724276	19	0.72371
2	0.724633	11	0.724203	20	0.723873
3	0.724716	12	0.724249	21	0.723819
4	0.723738	13	0.723154	22	0.723945
5	0.724714	14	0.724004	23	0.723782
6	0.724467	15	0.724051	24	0.723759
7	0.724533	16	0.723862	25	0.723975
8	0.724526	17	0.724141	26	0.723982
9	0.724561	18	0.723814	27	0.72379

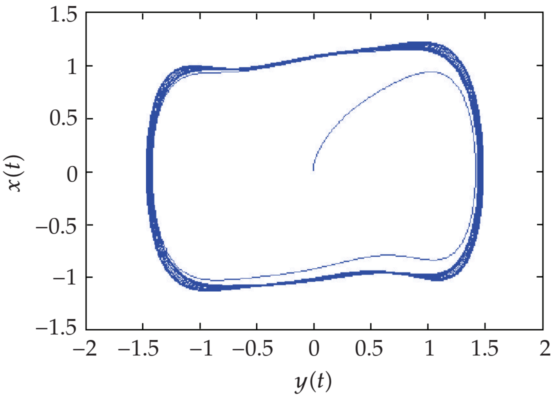


(a)

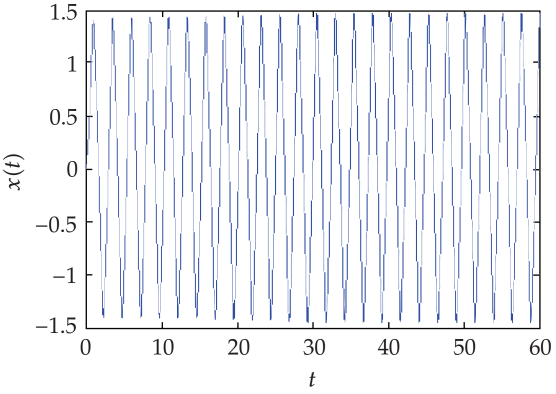
(b)

Figure 9: Sharp_1 phase plane and time domain.

Here, we meet a problem. When the computation precision of the threshold r is not appropriate, dynamic system 10 is not stable. In other words, perhaps one group data is chaotic and another group data is periodic state in all wear tools. Behavior of the dynamic system is changed with the threshold difference, see Figure 4. In order to decrease computation cost with Matlab, we fix a step size 0.1 in the region $r = [0, 1]$, the system trends



(a)



(b)

Figure 10: Wear_1 phase plane and time domain.

of dynamic behavior can be get roughly, this computing process spends about 40 minutes. In fact, the region $r = [0.7, 0.8]$ is the region from chaotic state to periodic state for Duffing-Holmes oscillator, no matter what any exterior weak periodic signals are merged into the system, the $r = [0.7, 0.8]$ can be used directly as ruler.

If we took the critical value of each group data as the threshold value, we would get 27 difference threshold values. However, we must get one threshold value for all 27 group data. For the reason, the range of the threshold value will be enlarged, that is, the threshold value will be decreased. We take the minimum threshold value in all 27 group sharp tool data. Using the law of the golden section in the region $r = [0.7, 0.8]$ for each group data of sharp tool, their critical value are calculated, see Table 4. Obviously, minimum is 0.723710.

When the amplitude of interior perturbation force r is equal to 0.723710, System 9 is critical state from chaotic to periodic, one of 27 groups data of sharp tool is merged to system 9, the system shows chaotic state, and when one of 27 groups of wear tool is merged to system 10, the system shows larger scale periodic state. For first group data, system 9 and system 10 are showed Figures 9 and 10.

The computation precision is six digits after decimal dot for the threshold determined accurately, it is enough sensitivity for distinguish wear tool or sharp tool. Of course, the

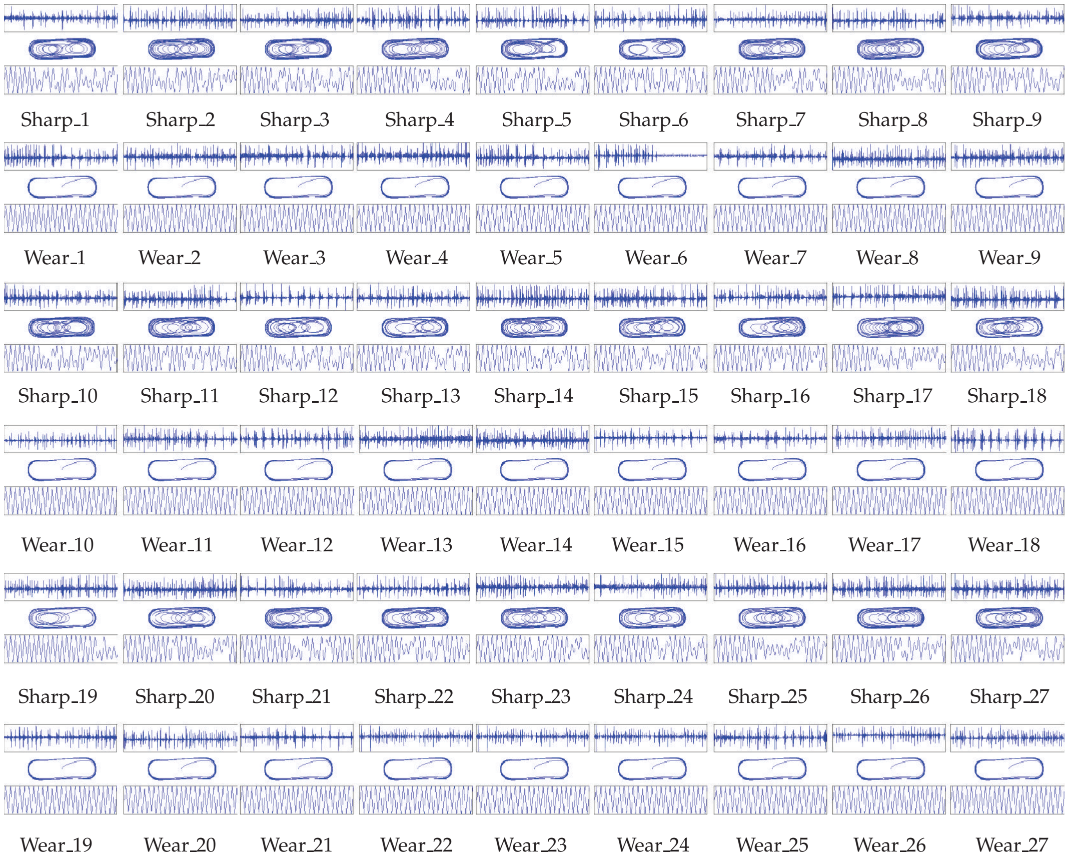


Figure 11: Map of time domain to be detected signal, phase plane, and state of time domain on system 9.

difference engineering problem may choose the difference computation precision for the threshold value in chaotic system model.

Finally, 0.723710 is a threshold value detected wear tool. All 54 group data sampled is merged to system 9, respectively, time domain map of each group data sampled, system phase plane, time domain map of system state; they are shown in Figure 11.

6. Conclusion

Currently, Duffing-Holmes oscillator is the area of most intense research activity for developing weak signal detected. The method which was described in this paper can be used as a valuable tool for the tool condition monitoring. In comparison to conventional weak signal detected, the advantages of tool wear detected based on Duffing-Holmes oscillator were shown. Compared to the Lyapunov exponents calculated determining the threshold of system chaotic critical state, the law of the golden section spends the less time and useful engineering meaning. The computation precision of the threshold can be calculated conventionally to satisfy the sensitivity of wear tool detected.

For the future development of the presented techniques in laboratory, several 10 approaches are to be tested. For example, relationship between the computation precision of the threshold and sensitivity of the chaotic critical state for difference engineering problem.

Since Runge-Kutta method of fourth order is one kind of approximate solution method for dynamic equation, a difference time-step size will impact the computation precision for the threshold value.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under the project Grant no. 60573125.

References

[1] S. Damodarasamy and S. Raman, "An inexpensive system for classifying tool wear states using pattern recognition," *Wear*, vol. 170, no. 2, pp. 149–160, 1993.

[2] S. Wanqing, Y. Jianguo, and Q. Chen, "Tool condition monitoring based on fractal and wavelet analysis by acoustic emission," in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA '07)*, vol. 4705 of *Lecture Notes in Computer Science*, pp. 469–479, Kuala Lumpur, Malaysia, August 2007.

[3] Y. Yao, X. Li, and Z. Yuan, "Tool wear detection with fuzzy classification and wavelet fuzzy neural network," *International Journal of Machine Tools and Manufacture*, vol. 39, no. 10, pp. 1525–1538, 1999.

[4] S.-S. Cho and K. Komvopoulos, "Correlation between acoustic emission and wear of multi-layer ceramic coated carbide tools," *Journal of Manufacturing Science and Engineering*, vol. 119, no. 2, pp. 238–246, 1997.

[5] N. Hu, X. Wen, and M. Chen, "Application of the Duffing chaotic oscillator model for early fault diagnosis-I. Basic theory," *International Journal of Plant Engineering and Management*, vol. 7, no. 2, pp. 67–75, 2006.

[6] Y. Li and B. Yang, "Chaotic system for the detection of periodic signals under the background of strong noise," *Chinese Science Bulletin*, vol. 48, no. 5, pp. 508–510, 2003.

[7] D. Liu, H. Ren, L. Song, and H. Li, "Weak signal detection based on chaotic oscillator," in *Proceedings of the 40th IAS Annual Meeting on Industry Applications Conference*, vol. 3, pp. 2054–2058, Hong Kong, October 2005.

[8] B. Le, Z. Liu, and T. Gu, "Chaotic oscillator and other techniques for detection of weak signals," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E88-A, no. 10, pp. 2699–2701, 2005.

[9] S. Zheng, H. Guo, Y. Li, B. Wang, and P. Zhang, "A new method for detecting line spectrum of ship-radiated noise using duffing oscillator," *Chinese Science Bulletin*, vol. 52, no. 14, pp. 1906–1912, 2007.

[10] J. C. Sprott, *Chaos and Time-Series Analysis*, Oxford University Press, New York, NY, USA, 2003.

[11] F. Grond and H. H. Diebner, "Local Lyapunov exponents for dissipative continuous systems," *Chaos, Solitons & Fractals*, vol. 23, no. 5, pp. 1809–1817, 2005.

[12] G. Wang, D. Chen, J. Lin, and X. Chen, "The application of chaotic oscillators to weak signal detection," *IEEE Transactions on Industrial Electronics*, vol. 46, no. 2, pp. 440–444, 1999.

[13] F. C. Moon, *Chaotic and Fractal Dynamics: An Introduction for Applied Scientists and Engineers*, A Wiley-Interscience Publication, John Wiley & Sons, New York, NY, USA, 1992.

[14] T. A. Bartler, *Lyapunov exponents and chaos investigation*, Doctoral dissertation, University of Cincinnati, Cincinnati, Ohio, USA, 1999.

[15] F. E. Udwadia and H. F. von Bremen, "An efficient and stable approach for computation of Lyapunov characteristic exponents of continuous dynamical systems," *Applied Mathematics and Computation*, vol. 121, no. 2-3, pp. 219–259, 2001.

[16] F. E. Udwadia and H. F. von Bremen, "Computation of Lyapunov characteristic exponents for continuous dynamical systems," *Zeitschrift für Angewandte Mathematik und Physik*, vol. 53, no. 1, pp. 123–146, 2002.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 807287, 25 pages
doi:10.1155/2008/807287

Research Article

Venturi Wet Gas Flow Modeling Based on Homogeneous and Separated Flow Theory

Fang Lide,[1] Zhang Tao,[2] and Xu Ying[2]

[1] *The Institute of Quality and Technology Supervising, Hebei University, Baoding 071051, China*
[2] *School of Electrical and Automation Engineering, Tianjin University, Weijin Road no. 92, Tianjin 300072, China*

Correspondence should be addressed to Fang Lide, leed_amy@yahoo.com.cn

Received 16 April 2008; Accepted 27 May 2008

Recommended by Cristian Toma

When Venturi meters are used in wet gas, the measured differential pressure is higher than it would be in gas phases flowing alone. This phenomenon is called over-reading. Eight famous over-reading correlations have been studied by many researchers under low- and high-pressure conditions, the conclusion is separated flow model and homogeneous flow model performing well both under high and low pressures. In this study, a new metering method is presented based on homogeneous and separated flow theory; the acceleration pressure drop and the friction pressure drop of Venturi under two-phase flow conditions are considered in new correlation, and its validity is verified through experiment. For low pressure, a new test program has been implemented in Tianjin University's low-pressure wet gas loop. For high pressure, the National Engineering Laboratory offered their reports on the web, so the coefficients of the new proposed correlation are fitted with all independent data both under high and low pressures. Finally, the applicability and errors of new correlation are analyzed.

1. Introduction

Wet gas metering has been described as a subset of multiphase flow measurement, where the volume of gas at actual measuring conditions is very high when compared to the volume of liquid in the flow stream. High-gas volume fraction has been defined in the range of 90–98% by different technical papers; more details are shown by Agar and Farchy [1]. Normally, these conditions need wet gas metering; for instance, some small or remote gas fields are processed together in common platform facilities, the individual unprocessed streams must be metered before mixing. In other circumstances, some gas meters may also be subjected to small amounts of liquid in the gas. This can happen to the gas output of a separator as a result of unexpected well conditions or liquid slugging.

Table 1: Result of high-pressure comparison.

Models	Root mean square error	Rank
De Leeuw	0.0211	1
Homogeneous	0.0237	2
Lin	0.0462	3
Murdork 1.5	0.0482	4
Murdork 1.26	0.0650	5
Chisholm	0.0710	6
Smith and Leang	0.1260	7

Two ways are employed to meter wet gas: one approach is to use a multiphase flow meter in wet gas, and the other approach is to use a standard dry gas meter applying corrections to the measurements based on knowledge of how this type of meter is affected by the presence of liquid in the gas stream. This method requires prior knowledge of the liquid flow, which has to be obtained through another means; more details were shown by Lupeau et al. [2].

As a mature single-phase flow measurement device, the Venturi meter has been successfully applied in a variety of industrial fields and scientific research. Just owing to its successful applications in single-phase flows, the Venturi meter can easily be considered for two-phase flow measurement. When Venturi meters are used in wet gas, the measured differential pressure is higher than it would be with the gas phase flowing alone. If uncorrected, this additional pressure drop will result in an over reading of the gas mass flow rate. More details were shown by Geng et al. [3].

Eight famous over-reading correlations have been studied in low- and high-pressure conditions [4–10]. In Steven's paper [10], an ISA Controls standard North Sea specification $6''$ Venturi meters with a 0.55 diameter ratio (or "beta") of 6 mm pressure tappings was the meter installed in National Engineering Laboratory (NEL) with pressure from 2 to 6 MPa and LM parameter from 0 to 0.3. NEL's engineer tested three 4-inch meters with different beta values (0.4, 0.60, 0.75) and tested over a range of pressures (1.5–6.0 MPa) gas densimetric Froude number (Fr_g), 0.5–5.5, and Lockhart-Martinelli parameter, X, 0–0.4 [11–13]. The results show that the liquid existence causes the meters to "over-read" the gas flow rate. This over reading is affected by the liquid fraction, gas velocity, pressure, and Venturi beta value. They predicted that some of the data seem to tend to a value slightly above unity, particularly at low X values. Furthermore, in 2002, Britton et al. did some tests in Colorado Engineering Experiment Station, Inc., Colo, USA, [14, 15] with pressure between 1.4–7.6 MPa and X values between 0–0.25. Their study also confirmed the over-reading existence in Venturi meters.

The result of high-pressure comparison is shown in Table 1 [10].

Under low pressure, eight correlations are compared with Tianjin University's low-pressure wet gas test facilities [16] (see Table 2).

The method of comparing the seven correlations performances was chosen to be by comparison of the root mean square error (defined as δ):

$$\delta = \sqrt{\frac{1}{N}\sum_{1}^{N}\left(\frac{\mathrm{OR}_{p(i)} - \mathrm{OR}_{e(i)}}{\mathrm{OR}_{e(i)}}\right)^2}, \tag{1.1}$$

Table 2: Result of low-pressure comparison.

Models	RMSE	Rank
Homogenous	0.11021	1
Steven	0.14787	2
De Leeuw	0.14854	3
Smith and Leang	0.18821	4
Chisholm	0.19597	5
Murdock1.5	0.20658	6
Lin	0.20742	7
Murdock	0.21078	8

where $\mathrm{OR}_{p(i)}$ is prediction over reading; $\mathrm{OR}_{e(i)}$ is experimentation over reading; N is data numbers.

Tables 1 and 2 show the models performance in low and high pressure. By De Leeuw model being based on separated flow assumption, more parameters have been considered so it performs well. Although the assumptions of homogeneous models are simple, it performs well at both low pressure and high pressure (see Steven's results), for wet gas, homogeneous models may be true to some extent. This means that wet gas flow structure holds homogeneous character and separation character. Therefore, a new correlation considering homogeneous and separation flow theory together could be better than the previous ones.

This paper proposed a new Venturi wet gas correlation based on homogenous and separate assumption. The acceleration pressure drop and the friction pressure drop of Venturi under two-phase flow conditions are considered in new correlation, and its validity is verified through experiment. Finally, the performance of the new proposed correlations is compared with the old eight correlations both under low and high pressure.

2. New Model Based on Homogeneous and Separated Flow Theory

2.1. Over-Reading Theory of Venturi Wet Gas Metering

When Venturi meters are used in wet gas the measured differential pressure is higher than it would be for the gas phase flowing alone. If uncorrected, this additional pressure drop will result in an over reading of the gas mass flow rate:

$$\mathrm{OR} = \frac{m_g'}{m_g}, \tag{2.1}$$

where m_g is the correct gas mass flow rate, m_g' is the apparent gas mass flow rate determined from the two-phase measured differential pressure ΔP_{tp}, ΔP_{tp} is the actual two-phase differential pressure between the upstream and throat tappings, and ΔP_g is the gas differential pressure between the upstream and throat tappings:

$$m_g = \frac{C\varepsilon A_T\sqrt{2\rho_g \Delta P_g}}{\sqrt{1-\beta^4}}, \tag{2.2}$$

$$m_g' = \frac{C\varepsilon A_T\sqrt{2\rho_g \Delta P_{\text{tp}}}}{\sqrt{1-\beta^4}}. \tag{2.3}$$

In (2.2) and (2.3), C is discharge coefficient, A_T is the area of the Venturi throat, ε is expansibility factor, ρ_g is gas density, and β is diameter ratio. In fact, the discharge coefficient C is variable under different flow conditions. Here, given that the discharge coefficient C is constant, and take into account the fact that different flow conditions only have effect on over reading, but not have effect on the discharge coefficient given C.

The real gas mass flow rate can been obtained by

$$m_g = \frac{m_g'}{\text{OR}}. \tag{2.4}$$

The homogeneous flow theory treats the two-phase flow as if it was a single-phase flow by using a homogeneous density expression ρ_{tp} which averages the phase densities so that the single-phase differential pressure meter equation can be used

$$\frac{1}{\rho_{\text{tp}}} = \frac{x}{\rho_g} + \frac{1-x}{\rho_l}, \tag{2.5}$$

where x is the mass quality, ρ_{tp} is the homogeneous density, and subscripts "l" and "g" are for liquid and gas, respectively.

With this models the gas mass flow rate of the two phase flow can be written as

$$m_g = x\frac{C\varepsilon A_T\sqrt{2\rho_{\text{tp}} \Delta P_{\text{tp}}}}{\sqrt{1-\beta^4}}. \tag{2.6}$$

Let (2.3) divide (2.6), then the homogeneous model gives

$$\begin{aligned} \text{OR}_h &= \frac{m_g'}{m_g} = \frac{C\varepsilon A_T\sqrt{2\rho_g \Delta P_{\text{tp}}}\Big/\sqrt{1-\beta^4}}{x\left(C\varepsilon A_T\sqrt{2\rho_{\text{tp}} \Delta P_{\text{tp}}}\Big/\sqrt{1-\beta^4}\right)}, \\ \text{OR}_h &= \frac{1}{x}\sqrt{\frac{\rho_g}{\rho_l} + \left(1 - \frac{\rho_g}{\rho_l}\right)x}. \end{aligned} \tag{2.7}$$

However, (2.6) is also an estimation function about gas mass flow rate; the real gas mass flow rate should be (2.2) and then (2.6) as the apparent gas mass flow rate will be more rational. So let (2.6) divide (2.2), the real over reading under the homogeneous flow theory is shown in the following form:

$$\begin{aligned} \mathrm{OR}_h &= \frac{m'_g}{m_g} \\ &= \frac{x \cdot \left(C\varepsilon A_T \sqrt{2\rho_{\mathrm{tp}} \Delta P_{\mathrm{tp}}} \Big/ \sqrt{1-\beta^4} \right)}{C\varepsilon A_T \sqrt{2\rho_g \Delta P_g} \Big/ \sqrt{1-\beta^4}} \\ &= x \cdot \sqrt{\frac{\rho_{\mathrm{tp}}}{\rho_g} \cdot \frac{\Delta P_{\mathrm{tp}}}{\Delta P_g}}. \end{aligned} \tag{2.8}$$

Equation (2.8) derived from homogeneous flow theory, if $\sqrt{\Delta P_{\mathrm{tp}}/\Delta P_g}$ derived from separation flow theory, the combination of homogeneous and separation flow theory is implemented.

Separated flow theory takes into account the fact that the two phases can have differing properties and different velocities. Separate equations of continuity, momentum, and energy are written for each phase, and these six equations are solved simultaneously, together with rate equations which describe how the phases interact with each other and with the walls of duct. In the simplest version, only one parameter, such as velocity, is allowed to differ for the two phases while conservation equations are only written for the combined flow.

Equation (2.9) shows the momentum function of one dimension two-phase flow based on separated flow assumption. The pressure drop of fluids in the pipe come from three parts, the first is friction; the second is gravitation; the third is acceleration [17–21]:

$$\begin{aligned} -\frac{dP}{dz} &= \frac{\tau_0 U}{A} + [\rho_g \alpha + \rho_\mathrm{l}(1-\alpha)] g \sin\theta \\ &\quad + \frac{1}{A}\frac{d}{dz}\left\{ AG^2 \left[\frac{(1-x)^2}{\rho_l(1-\alpha)} + \frac{x^2}{\rho_g \alpha} \right] \right\}, \end{aligned} \tag{2.9}$$

$$-\frac{dP}{dz} = \frac{dP_f}{dz} + \frac{dP_g}{dz} + \frac{dP_a}{dz}, \tag{2.10}$$

where τ_0 is friction force, U is perimeter of pipe, α is void fraction, G is mass velocity of mixture, dP_f/dz is pressure drop caused by friction, dP_g/dz is pressure drop caused by gravitation, dP_a/dz is pressure drop caused by acceleration.

2.2. The Friction Pressure Drop of Venturi Under Two-Phase Flow Condition

For single-phase flow in straight pipe, the friction pressure drop can be calculated with

$$\frac{dP_f}{dz} = \frac{\lambda}{d} \cdot \frac{\rho u^2}{2}, \tag{2.11}$$

where λ is the friction factor; d is the pipe diameter, u is the velocity.

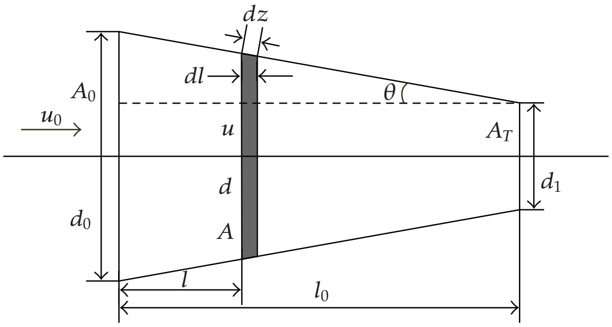


Figure 1: Sketch of Venturi conical convergent region.

Given λ is constant in conical convergent of Venturi, the fluid velocity in the straight pipe keep unchanged, d_0 is diameter of straight pipe, A_0 is cross-section of straight pipe, d_1 is diameter of Venturi throat, l_0 is the length of conical convergent, θ is convergent angle. The schematic of Venturi conical convergent part is shown in Figure 1.

Analyzing an infinitesimal dl given d is diameter of the analyzing part, A is cross-section, l is the distance from Venturi inlet to infinitesimal dl make integral to (2.11):

$$\Delta P_f = \int_0^{l_0} \frac{\lambda}{d} \cdot \frac{\rho u^2}{2}\, dz, \tag{2.12}$$

$$\Delta P_f = \frac{\lambda \rho}{2} \int_0^{l_0} \frac{1}{d} \cdot u^2\, dz. \tag{2.13}$$

Multiply d_0 to (2.13) in two sides:

$$\Delta P_f = \frac{\lambda \rho}{2 d_0} \int_0^{l_0} \frac{d_0}{d} \cdot u^2\, dz. \tag{2.14}$$

From continuity equation,

$$u = \frac{A_0}{A} u_0, \qquad \frac{A_0}{A} = \left(\frac{d_0}{d}\right)^2. \tag{2.15}$$

Substitute (2.15) into (2.14):

$$\Delta P_f = \frac{\lambda \rho u_0^2}{2 d_0} \int_0^{l_0} \left(\frac{d_0}{d}\right)^5 dz. \tag{2.16}$$

According to geometrical relationship showed in Figure 1,

$$dz = \frac{dl}{\cos\theta}, \tag{2.17}$$

$$\frac{l}{l_0} = \frac{d_0 - d}{d_0 - d_1}, \tag{2.18}$$

$$\Longrightarrow \frac{d_0}{d} = \frac{l_0}{l_0 - l(1 - d_1/d_0)}. \tag{2.19}$$

Let β be diameter ratio of Venturi, then

$$\beta = \frac{d_1}{d_0}. \tag{2.20}$$

Substitute (2.17), (2.19), and (2.20) into (2.16):

$$\Delta P_f = \frac{\lambda \rho u_0^2 l_0^5}{2 d_0 \cos\theta} \int_0^{l_0} \left(\frac{1}{l_0 - l(1-\beta)} \right)^5 dl, \tag{2.21}$$

$$\Delta P_f = \frac{\lambda \rho u_0^2 l_0^5}{8 d_0 \cos\theta (1-\beta)} \left(\frac{1}{l_0 - l(1-\beta)} \right)^4 \Bigg|_0^{l_0}, \tag{2.22}$$

$$\Delta P_f = \frac{(1+\beta)\cdot(1+\beta^2)}{\beta^4} \cdot \frac{1}{4\cos\theta} \cdot \frac{\lambda}{d_0} \cdot \frac{\rho u_0^2}{2} l_0. \tag{2.23a}$$

Equation (2.23a) shows that the friction pressure drop is affected by diameter ratio, convergent angle, convergent length, inlet diameter, and inlet velocity.

In a constant section pipe with l_0 length, the friction pressure drop is

$$\Delta P_{f_{l0}} = \frac{\lambda l_0}{d_0} \cdot \frac{\rho u_0^2}{2}. \tag{2.23b}$$

Equation (2.23a) that is divided by (2.23b) is

$$K_f = \frac{(1+\beta)\cdot(1+\beta^2)}{\beta^4} \cdot \frac{1}{4\cos\theta}. \tag{2.24}$$

Equation (2.24) shows that the ratio K_f is a function of diameter ratio and convergent angle. For a definite Venturi, K_f is constant.

As for gas liquid two-phase flow, (2.23a) and (2.23b) changes into

$$\Delta P_f = K_f \cdot \frac{\lambda l_0}{d_0} \cdot \frac{\alpha \rho_g u_g^2 + (1-\alpha)\rho_l u_l^2}{2}. \tag{2.25}$$

When the pipe is full of gas ($\alpha = 1$) or liquid ($\alpha = 0$), (2.25) changes to (2.23a).

From gas liquid two-phase flow continuity equation,

$$\begin{aligned} xGA &= A_g u_g \rho_g, \\ (1-x)GA &= A_l u_l \rho_l. \end{aligned} \tag{2.26}$$

Consider the definition of void fraction,

$$\begin{aligned} \frac{x}{\alpha} G &= u_g \rho_g, \\ \frac{(1-x)}{(1-\alpha)} G &= u_l \rho_l, \end{aligned} \tag{2.27}$$

$$G = \frac{m}{A} = \alpha \rho_g u_g + (1-\alpha)\rho_l u_l \tag{2.28}$$

which defined S as slip ratio, that is, gas and liquid real velocity ratio combine (2.26) and (2.27):

$$\frac{1}{\alpha} = 1 + s\frac{1-x}{x} \cdot \frac{\rho_g}{\rho_l}. \tag{2.29}$$

Substitute (2.26) and (2.27) into (2.25):

$$\Delta P_f = K_f \cdot \frac{\lambda l_0}{d_0} \cdot \frac{G^2}{2} \cdot \frac{1}{\rho_l}\left[\frac{x^2}{\alpha} \cdot \frac{\rho_l}{\rho_g} + \frac{(1-x)^2}{1-\alpha}\right]. \tag{2.30}$$

When the pipe is full of gas,

$$\Delta P_{\mathrm{fg}} = K_f \cdot \frac{\lambda_g l_0}{d_0} \cdot \frac{G^2}{2} \cdot \frac{x^2}{\rho_g}. \tag{2.31}$$

When the pipe is full of liquid,

$$\Delta P_{\mathrm{fl}} = K_f \cdot \frac{\lambda_l l_0}{d_0} \cdot \frac{G^2}{2} \cdot \frac{(1-x)^2}{\rho_l}. \tag{2.32}$$

Let $\lambda = \lambda_g = \lambda_l$, define X_f as

$$X_f = \sqrt{\frac{\Delta P_{fl}}{\Delta P_{fg}}} = \left(\frac{1-x}{x}\right)\sqrt{\frac{\rho_g}{\rho_l}}. \tag{2.33}$$

Equation (2.30) divided by (2.31) is

$$\frac{\Delta P_f}{\Delta P_{fg}} = \frac{1}{\alpha} + \frac{(1-x)^2}{x^2}\frac{\rho_g}{\rho_l} \cdot \frac{1}{1-\alpha}. \tag{2.34}$$

Substitute (2.29) into (2.34):

$$\frac{\Delta P_f}{\Delta P_{fg}} = 1 + C_f X_f + X_f^2, \tag{2.35}$$

where

$$C_f = \frac{1}{s}\sqrt{\frac{\rho_l}{\rho_g}} + s\sqrt{\frac{\rho_g}{\rho_l}}. \tag{2.36}$$

2.3. The Acceleration Pressure Drop of Venturi Under Two-Phase Flow Condition

According to (2.9), the acceleration pressure drop is

$$\frac{dP_a}{dz} = \frac{1}{A}\frac{d}{dz}\left\{AG^2\left[\frac{(1-x)^2}{\rho_l(1-\alpha)} + \frac{x^2}{\rho_g\alpha}\right]\right\}, \tag{2.37}$$

$$\Delta P_a = \int dP_a = \int_{A_T}^{A_0} \frac{1}{A}\, d\left\{AG^2\left[\frac{(1-x)^2}{\rho_l(1-\alpha)} + \frac{x^2}{\rho_g\alpha}\right]\right\}. \tag{2.38}$$

Given the fluid is incompressible, the void fraction α is constant in the Venturi throat. Integrate (2.38):

$$\Delta P_a = G^2 \left[\frac{(1-x)^2}{\rho_l (1-\alpha)} + \frac{x^2}{\rho_g \alpha} \right] \cdot \ln \frac{A_0}{A_T}. \tag{2.39}$$

When the gas was flowing alone in the pipe, the pressure drop can be expressed as

$$\Delta P_{ag} = G^2 \frac{x^2}{\rho_g} \cdot \ln \frac{A_0}{A_T}. \tag{2.40}$$

The similar equation for the liquid phase is

$$\Delta P_{al} = \frac{G^2 (1-x)^2}{\rho_l} \cdot \ln \frac{A_0}{A_T}. \tag{2.41}$$

Define X_a:

$$X_a = \sqrt{\frac{\Delta P_{al}}{\Delta P_{ag}}} = \frac{(1-x)}{x} \cdot \sqrt{\frac{\rho_g}{\rho_l}}. \tag{2.42}$$

Equation (2.39) divided by (2.40) is

$$\frac{\Delta P_a}{\Delta P_{ag}} = \frac{(1-x)^2}{x^2} \cdot \frac{\rho_g}{\rho_l} \cdot \frac{1}{1-\alpha} + \frac{1}{\alpha}. \tag{2.43}$$

Substitute (2.29) and (2.42) into (2.34):

$$\frac{\Delta P_a}{\Delta P_{ag}} = 1 + C_a \cdot X_a + X_a^2, \tag{2.44}$$

where C_a is expressed as

$$C_a = \frac{1}{s} \cdot \sqrt{\frac{\rho_l}{\rho_g}} + s \cdot \sqrt{\frac{\rho_g}{\rho_l}}. \tag{2.45}$$

Compare (2.33) with (2.42), it is obvious that X_f is the same as X_a.

Also, compared (2.33) with (2.42), C_f is equal to C_a.

And then, (2.44) is equal to

$$\frac{\Delta P_f}{\Delta P_{fg}} = \frac{\Delta P_a}{\Delta P_{ag}} = 1 + C_g \cdot X + X^2, \tag{2.46}$$

where

$$C_g = C_a = C_f = \frac{1}{s} \cdot \sqrt{\frac{\rho_l}{\rho_g}} + s \cdot \sqrt{\frac{\rho_g}{\rho_l}}, \tag{2.47}$$

$$X = X_a = X_f = \frac{(1-x)}{x} \cdot \sqrt{\frac{\rho_g}{\rho_l}}. \tag{2.48}$$

Equation (2.46) notes that the ratio of two-phase and single-phase friction pressure drop is equal to the ratio of two- phase and single-phase acceleration pressure drop.

2.4. The Total Pressure Drop of Venturi Under Two-Phase Flow Condition

For a horizontal mounted Venturi, gravitation pressure drop can be ignored. The total pressure drop is

$$\Delta P_{\text{tp}} = \Delta P_f + \Delta P_a. \tag{2.49}$$

The total pressure drop of Venturi under single-phase flow condition is

$$\Delta P_g = \Delta P_{fg} + \Delta P_{ag}. \tag{2.50}$$

Divide (2.49) by (2.50):

$$\frac{\Delta P_{\text{tp}}}{\Delta P_g} = \frac{\Delta P_f + \Delta P_a}{\Delta P_{fg} + \Delta P_{ag}}. \tag{2.51}$$

According to (2.46) and geometric axiom,

$$\frac{\Delta P_f}{\Delta P_{fg}} = \frac{\Delta P_a}{\Delta P_{ag}} = \frac{\Delta P_f + \Delta P_a}{\Delta P_{fg} + \Delta P_{ag}}. \tag{2.52}$$

Combine (2.51) and (2.52):

$$\frac{\Delta P_{\text{tp}}}{\Delta P_g} = \frac{\Delta P_f}{\Delta P_{fg}} = \frac{\Delta P_a}{\Delta P_{ag}} = 1 + C_g \cdot X + X^2. \tag{2.53}$$

So the model combined homogeneous and separation flow theory can be expressed as (2.55). Call this correlation as *H-S* model:

$$\begin{aligned} \text{OR}_{\text{H-S}} &= x \cdot \sqrt{\frac{\rho_{\text{tp}}}{\rho_g} \cdot \frac{\Delta P_{\text{tp}}}{\Delta P_g}} \\ &= x \cdot \sqrt{\frac{\rho_l}{\rho_l x + \rho_g (1-x)}} \cdot \sqrt{1 + C_g {\cdot} X + X^2} \end{aligned} \tag{2.54}$$

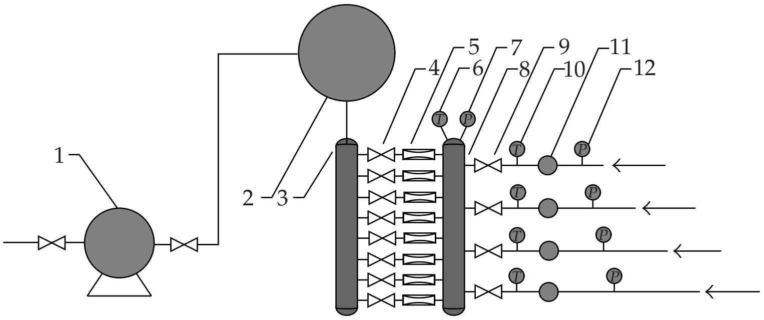
$$= \sqrt{x \cdot \frac{1 + C_g \cdot X + X^2}{1 + \sqrt{\rho_g / \rho_l}\, X}}. \tag{2.55}$$

Equation (2.55) shows that C_g is an effect factor to OR, it must be known first when (2.55) is used. However, slip ratio S is contained in C_g equation, and slip ratio is hard to be determined accurately, so it needs to fit a correlation with experiment.

3. Dry Gas Calibration and Wet Gas Tests

3.1. Dry Gas Calibration

Three venture meters are calibrated in TJU critical sonic nozzle flow calibration facility; see Figure 2.



1. Vacuum pump
2. Air tank
3. Collecting pipe
4. Switch valve
5. Critical sonic nozzle
6. Temperature
7. Manometer
8. Stagnation tank
9. Control valve
10. Temperature
11. Flow meter
12. Manometer

Figure 2: Schematic diagram of TJU critical sonic nozzle flow calibration facility.

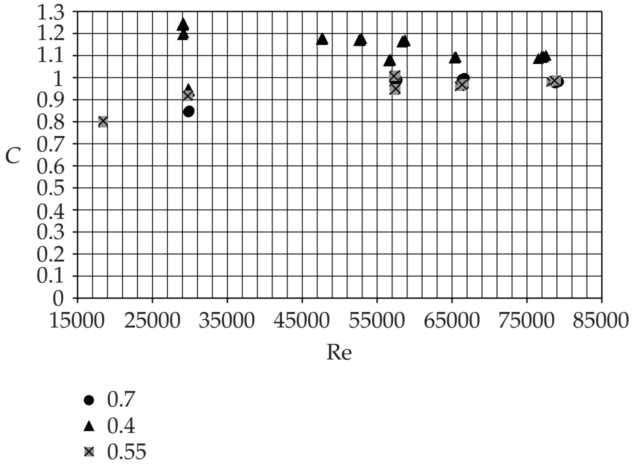


Figure 3: Discharge coefficient of Venturi tube in single phase flow.

The facility has eleven sonic nozzles of different discharge coefficient, and the calibration range varies from 2.50 to 660 m^3/h with a step of 2 m^3/h. The maximum calibrated flow rate is about 380 m^3/h due to the beta ratio and pipe diameter. At the same time, the TJU multiphase flow loop also has the calibration function. So the dry gas calibration for three Venturis was done in both. The test data from the two facilities show the same results. Figure 3 shows the calibration coefficient C with different diameter ratio. When the Reynolds number is higher than 1×10^5, the value of coefficient is in accord with the standard discharge coefficient for flows with Reynolds numbers less than one million [22].

Fit the coefficient C in different diameter ratio, the parameters listed in table 3:

$$C = P_1 + P_2 \cdot \mathrm{Re} + P_3 \cdot \mathrm{Re}^2 + P_4 \cdot \mathrm{Re}^3 + P_5 \cdot \mathrm{Re}^4 + P_6 \cdot \mathrm{Re}^5. \tag{3.1}$$

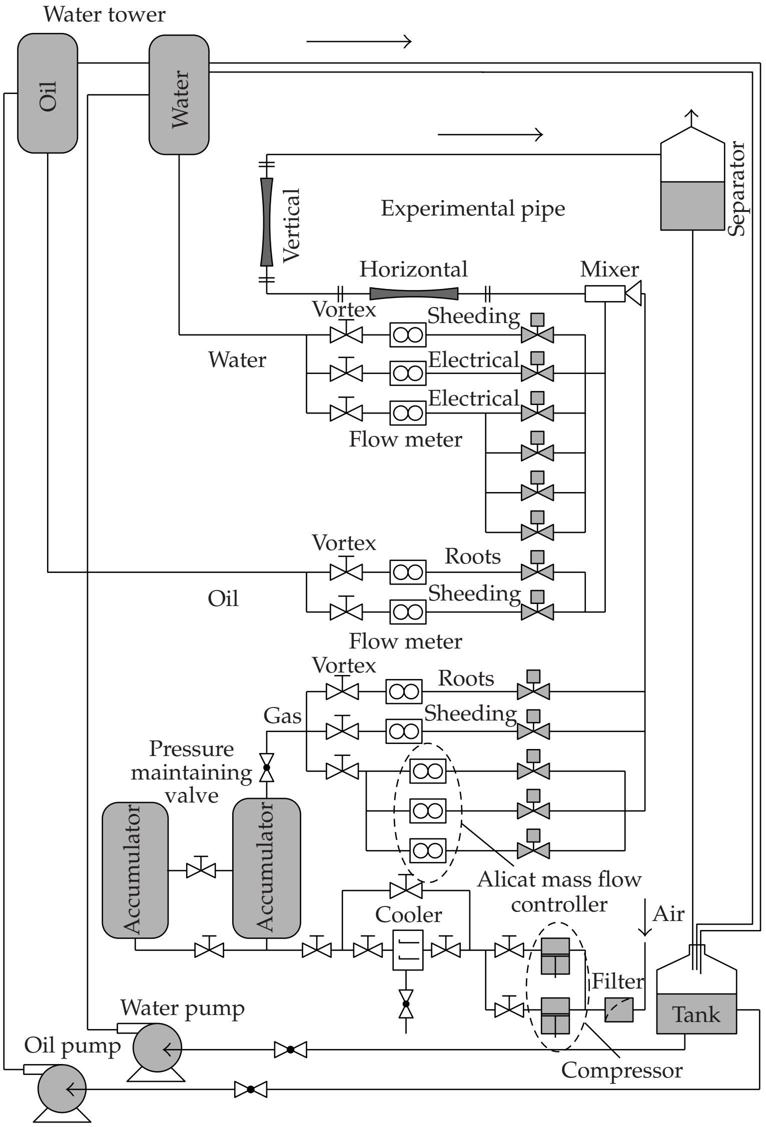


Figure 4: Schematic diagram of TJU multiphase flow loop.

3.2. Test System and Experimental Procedures

The tests were conducted on TJU multiphase flow loop at pressures from 0.15 MPa to 0.25 MPa across a range of gas velocities and liquid fractions. TJU's low-pressure wet gas test facilities are a fully automatic control and functional complete system, which is not only a multiphase flow experiment system, but also a multiphase flow meter calibration system. As an experiment system, the test can be conducted in a horizontal pipe, vertical pipe and 0–90°lean pipe; as a calibration system, the test meter can be calibrated in standard meter method. Figure 4 shows schematic diagram of TJU multiphase flow loop.

Thess facilities have six components, named as medium source, measurement pipe, horizontal pipe, vertical pipe, 0–90°lean pipe and computer control system.

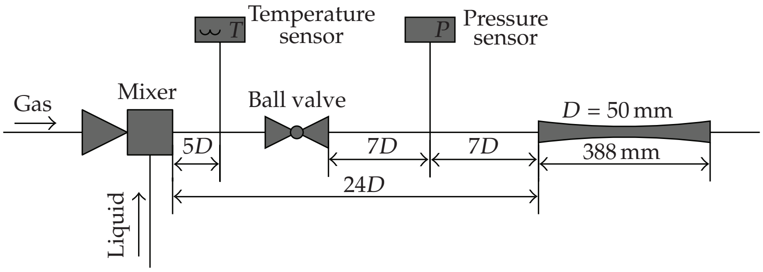


Figure 5: Horizontal experiment pipe.

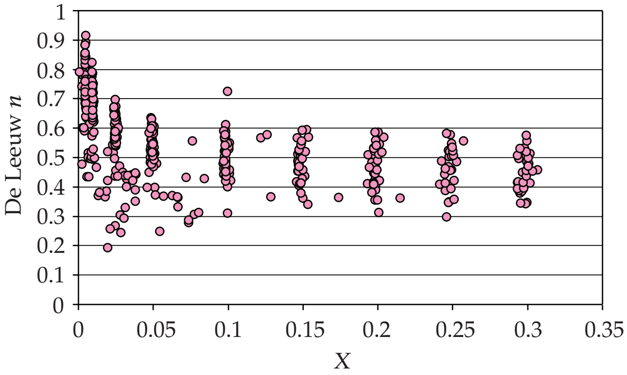


Figure 6: Lockhart-Martinelli parameter X effect on n of De Leeuw model.

Gas medium is compression air, and two compressors provide dynamic force, the compressor air is passing through cooling and drying unit which access to two 12 m^3 accumulator tanks; the accumulator tanks and pressure maintaining valve can hold a stable pressure 0–0.8 MPa for the test. The liquids used in test are water (oil or oil and water mixture also can be used) and a water pump pushes the water to a 30-meter-high water tower, which can hold a stable pressure for liquid.

In standard meter calibration system, gas calibration system has five paths; three of them are low-flow channels metering with three mass flow controllers made in America by Alicat scientific company, Ariz, USA, the lowest flux is 10 l/min, the other two paths are middle and large flow channels metering with a Roots type flow meter and a vortex flow meter. All temperature and pressure measurements use traceable calibrated instrumentation for gas temperature and pressure compensation.

Liquid calibration system has six paths: four of them are low-flow channels metering with an electrical flow meter made in Germany combined by four magnet valves, the lowest flux is 0.01 m^3/h, the other two paths are middle and large flow channels metering with a electrical flow meter and a vortex flow meter. See parameters of the standard meter in Table 4.

Gas and liquid calibrate through standard meter access to mixer, and then go through the experimental pipe. There are two paths in experimental pipes, one is made in rustless steel, the other is made in organic glass, their diameter is 50 mm, and a cutoff valve which can adjust the pressure is installed at outlet of the pipe.

Figure 5 shows horizontal experiment pipe, which includes mixer, temperature sensor, straight lengths, pressure sensor, and Venturi tube.

Table 3: Parameters value.

	0.4048	0.55	0.7
P_1	17.88251	−0.58867	6.06974
P_2	−0.00074	0.00006	−0.00016
P_3	$1.283E-8$	$-9.7022E-10$	$1.88E-9$
P_4	$-1.0944E-13$	$8.4168E-15$	$-1.0733E-14$
P_5	$4.5672E-19$	$-3.5094E-20$	$2.9549E-20$
P_6	$-7.4476E-25$	$5.5411E-26$	$-3.1419E-26$

Table 4: Parameters of the standard sensor.

Phases	Range (m^3/h)	Accuracy
Water	0.01 ~ 3.0	±0.2%
	0.75 ~ 19	±1.0%
	1.7 ~ 43	±0.5%
Air	0 ~ 6.0	±0.8%
	0.15 ~ 17	±3.0%
	6.5 ~ 130	±1.5%
Oil	0.02 ~ 2.5	1.0%
	0.75 ~ 19	1.0%

Table 5: Required straight lengths for classical Venturi tubes with a machined convergent section.

Diameter ratio	Straight length (D)
0.40	8
0.50	8
0.60	10
0.70	10
0.75	18

According to ISO 5167-1, 4 : 2003 [23, 24], a classical Venturi tube with a machined convergent section, straight lengths and diameter ratio must accord with Table 5.

In this test, three Venturi tubes with β values of 0.4048, 0.55, and 0.70 have been produced, the length of Venturi tubes is 388 mm, diameter is 50 mm, the length of cylindrical throat is 20 mm, conical convergent angle is 21°, conical divergent angle is 15°, diameter of pressure tappings is 4 mm, the pipe wall roughness is 0.06 mm, and stainless steel flange is used in connecting. 1151 differential pressure transducers were made in Rosemont company, Colo, USA, the uncertainty of whole equipment is 2.5‰.

The test data are collected and saved as Microsoft Excel file automatically (see experimental parameters in Table 6).

The flow pattern of the test included annular and drop-annular, where Fr_g is gas Froude number:

$$\mathrm{Fr}_g = \frac{v_g}{\sqrt{gD}}\sqrt{\frac{\rho_g}{\rho_l - \rho_g}}. \tag{3.2}$$

v_g is superficial velocity of the gas phase: $v_g = m_g/(\rho_g A)$.

Table 6: Experimental parameters.

β	P (MPa)	Fr_g	X
0.4048	0.15	0.8 ~ 1.5	0.0022 ~ 0.0338
	0.20	1.0 ~ 1.88	0.0022 ~ 0.0472
	0.25	0.67 ~ 1.81	0.0022 ~ 0.0495
0.55	0.15	1.04 ~ 1.78	0.0022 ~ 0.2572
	0.20	1.09 ~ 1.85	0.0022 ~ 0.3431
	0.25	0.92 ~ 1.73	0.0022 ~ 0.3514
0.7	0.15	1.04 ~ 2.0	0.0024 ~ 0.0480
	0.20	1.08 ~ 2.0	0.0025 ~ 0.0525
	0.25	0.87 ~ 1.66	0.0027 ~ 0.0576

Table 7: Fit exponent n with all data.

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	K
1.29203	−0.17161	0.12618	−0.01884	0.30196	0.05205	−0.07122	0.0259	0.78105

4. Model Parameters Determining and Error Analyzing

The coefficient C_g can be calculated by experimental data. On TJU multiphase flow system, the real gas, liquid mass flow rateand gas, liquid density can be determined by standard sensor. The gas mass fraction is known parameter. The Lockhart-Martinelli parameter can be obtained by (2.33). The over reading can be calculated with (2.1) and (2.3). Therefore, the coefficient C_g can be calculated by (2.55) (H-S model). The study shows that coefficient C_g decreases with increasing Lockhart-Martinelli parameter X, decreases with increasing pressure P, decreases with increasing diameter ratio β, increases with increasing Gas Froude number Fr_g,and increases with increasing gas liquid quality ratio $x/(1-x)$.

Equation (2.47) can be expressed as

$$C_g = f\left(s, \sqrt{\frac{\rho_g}{\rho_l}}\right). \tag{4.1}$$

Equation (4.1) shows that the gas-liquid quality ratio $x/(1-x)$ contains the same parameter with coefficient C_g:

$$\frac{m_g}{m_l} = \frac{x}{1-x} = S \cdot \frac{\rho_g}{\rho_l} \cdot \frac{\alpha}{1-\alpha}. \tag{4.2}$$

Combining (3.2) and (4.1) can gain

$$C_g = f\left(\frac{x}{1-x}, \frac{\alpha}{1-\alpha}, \sqrt{\frac{\rho_g}{\rho_l}}\right). \tag{4.3}$$

De Leeuw model considers the coefficient C_g as a function of gas-liquid density ratio and gas Froude number:

$$C_{\text{De Leeuw}} = \left(\frac{\rho_l}{\rho_g}\right)^n + \left(\frac{\rho_g}{\rho_l}\right)^n, \quad n = \begin{cases} 0.41 & 0.5 \leq \mathrm{Fr}_g \leq 1.5, \\ 0.606\left(1 - e^{-0.746\,\mathrm{Fr}_g}\right) & \mathrm{Fr}_g \geq 1.5. \end{cases} \tag{4.4}$$

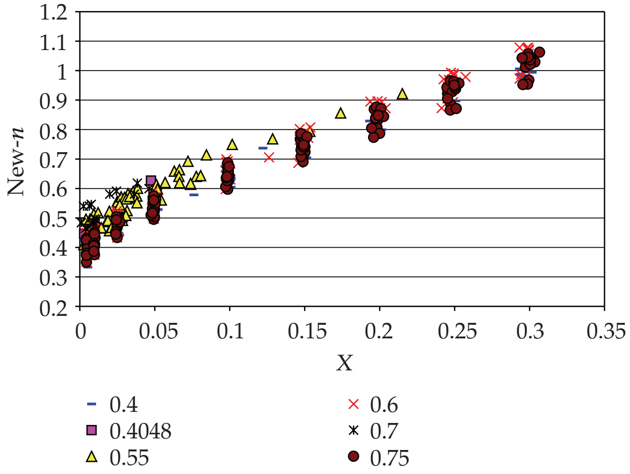


Figure 7: Lockhart-Martinelli parameter X effect to n.

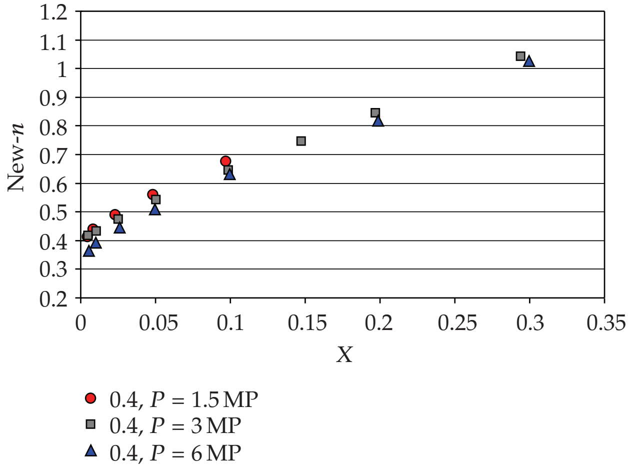


Figure 8: Pressure P effect on coefficient n under same gas Froude number.

However, inherited the form of the coefficient C_g of De Leeuw model's, and using gas liquid density ratio as base of exponential function, the exponent n is a severe nonlinear curve with other parameters such as Lockhart-Martinelli parameter X, or gas Froude number (see Figure 6).

Research found that using gas liquid volume ratio (gas liquid mass ratio divided by gas liquid density ratio) as a base of exponential function C_g in H-S model, the exponent n almost linear increases with increasing Lockhart-Martinelli parameter X, it can be seen as Figure 7, so defined the coefficient C_g of the H-S model as

$$C_{\text{H-S}} = \left(\frac{x/(1-x)}{\rho_g/\rho_l}\right)^n + \left(\frac{\rho_g/\rho_l}{x/(1-x)}\right)^n. \tag{4.5}$$

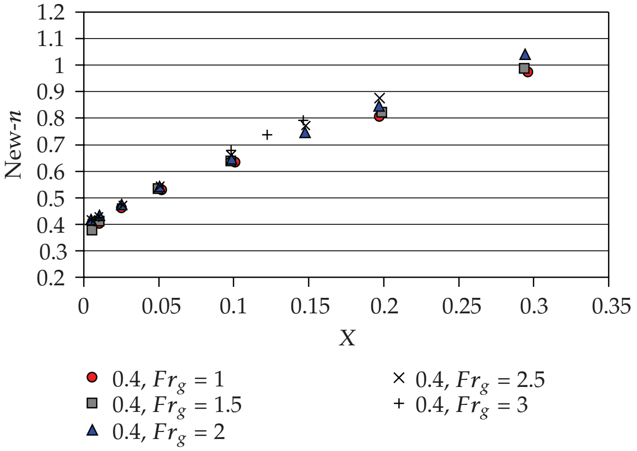


Figure 9: Gas Froude number Fr_g effect on coefficient n under same pressure.

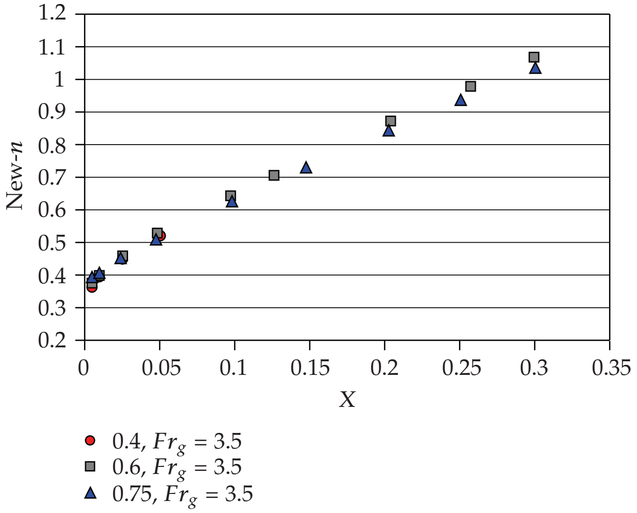


Figure 10: Diameter ratio effect on coefficient n under 6 MPa.

In fact, gas liquid mass ratio divided by gas liquid density ratio is equal to gas liquid volume ratio:

$$C_{\text{H-S}} = \left(\frac{\varphi}{1-\varphi}\right)^n + \left(\frac{1-\varphi}{\varphi}\right)^n,$$
$$n = f\left(\beta, P\left(\text{OR } \frac{\rho_g}{\rho_l}\right), \mathrm{Fr}_g, \mathrm{X}, \dots\right), \tag{4.6}$$

where φ is gas volume fraction.

Next, a correlation of exponent n with other parameters will be approached.

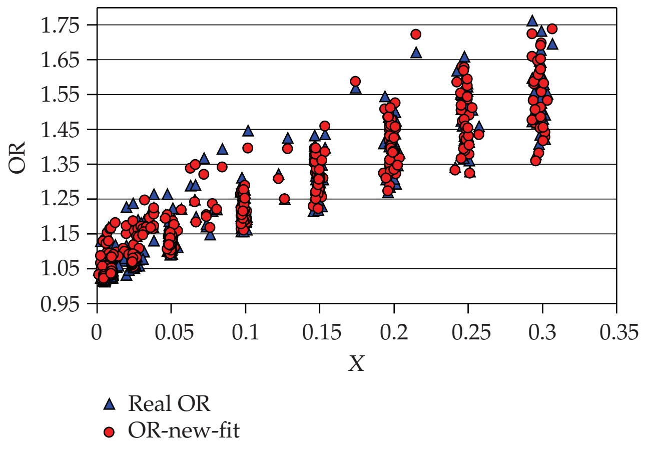


Figure 11: Comparison of H-S prediction OR and experimental OR.

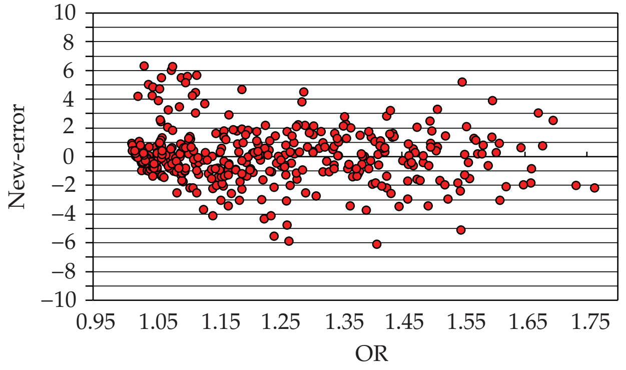


Figure 12: The prediction error of H-S model.

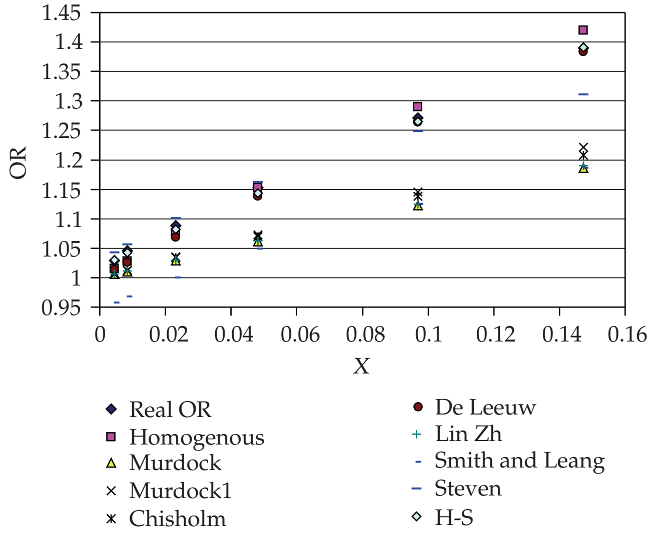


Figure 13: Comparison of models under $\beta = 0.4$, $P = 1.5\,\text{MPa}$, and $\text{Fr}_g = 2$.

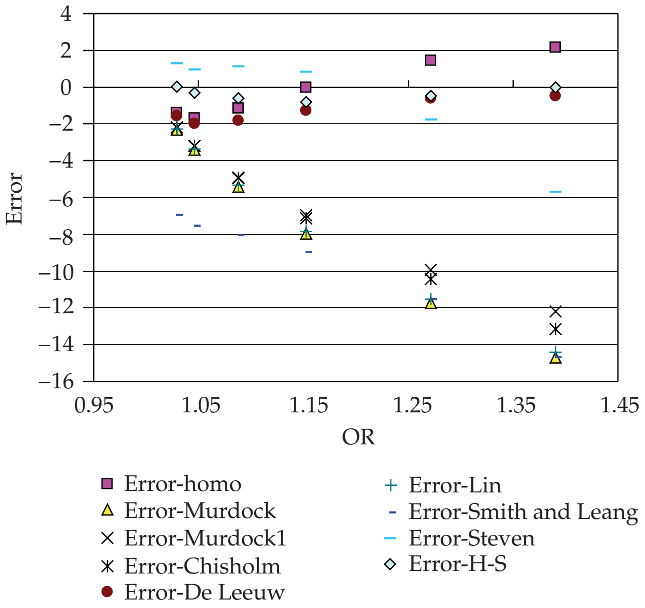


Figure 14: Comparison errors of models under $\beta = 0.4$, $P = 1.5\,\text{MPa}$, and $\text{Fr}_g = 2$.

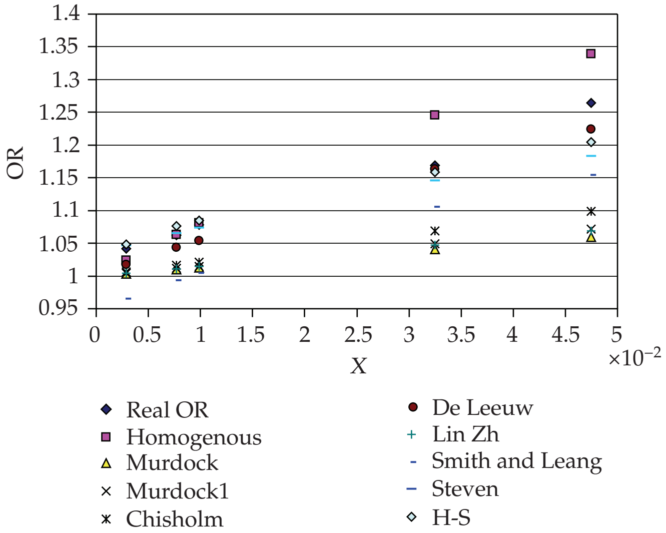


Figure 15: Comparison of models under $\beta = 0.4048$, $P = 0.20\,\text{MPa}$, and $\text{Fr}_g = 1.5$.

4.1. Effect of Parameters to Exponent n of H-S Model

Figure 7 shows the effect of Lockhart-Martinelli parameter X to n, and exponent n almost linearly increases with increasing Lockhart-Martinelli parameter X. Figure 8 shows the effect of pressure to n, apparently, exponent n decreases with the increasing pressure. Figure 9 shows the effect of gas Froude number to n, seemingly, n increases with the increasing of gas Froude number. Figure 10 shows the effect of diameter ratio to n, and n decreases with the increasing diameter ratio.

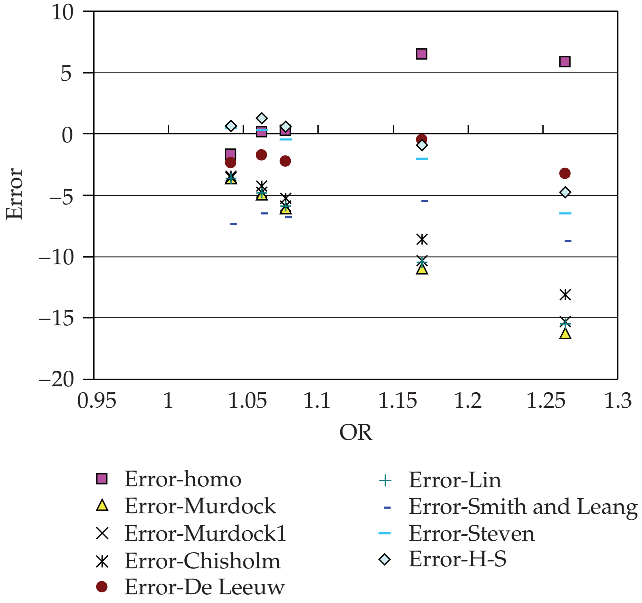


Figure 16: Comparison errors of models under $\beta = 0.4048$, $P = 0.20\,\text{MPa}$, and $\text{Fr}_g = 1.5$.

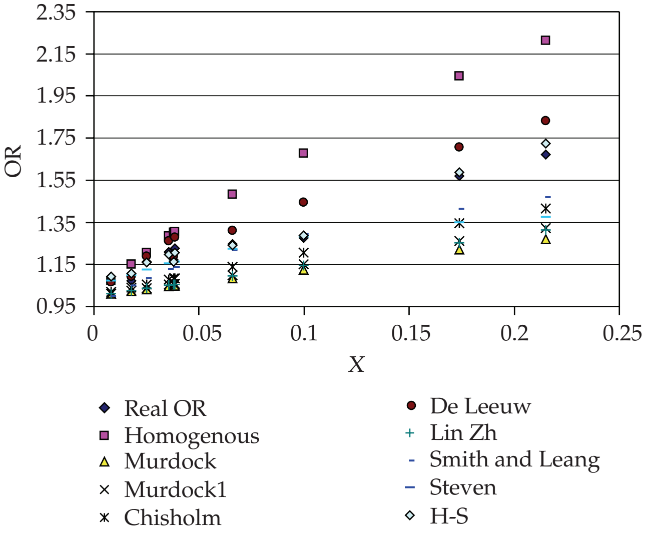


Figure 17: Comparison of models under $\beta = 0.55$, $P = 0.15\,\text{MPa}$, and $\text{Fr}_g = 2$.

4.2. Fitting Exponent n of H-S Model

According to the results of these figures, n varied linearly with Lockhart-Martinelli parameter X, and with the rate of curves effect by diameter ratio, pressure, and Gas Froude number. So the experiment correlation of coefficient n should take the Lockhart-Martinelli parameter X as a key independent variable, and pressure P (or gas liquid density ratio), diameter ratio β,

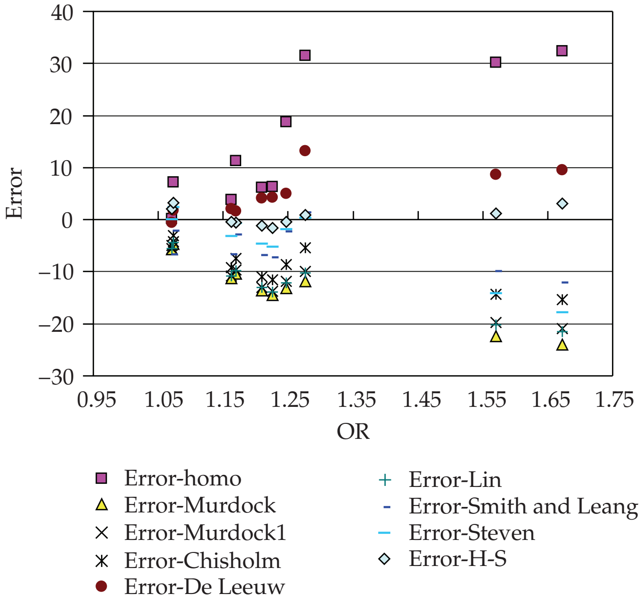


Figure 18: Comparison errors of models under $\beta = 0.55$, $P = 0.15\,\text{MPa}$, and $\text{Fr}_g = 2$.

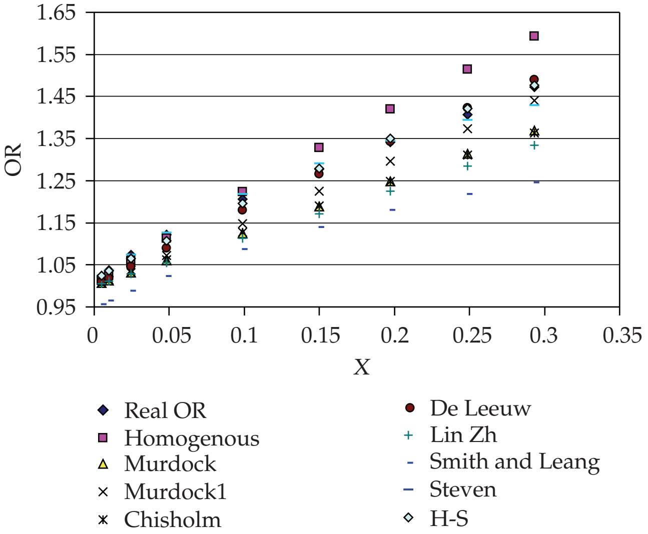


Figure 19: Comparison of models under $\beta = 0.60$, $P = 3\,\text{MPa}$, and $\text{Fr}_g = 1.5$.

Gas Froude number Fr_g as auxiliary variable. Exponent n can be defined as

$$n = A + B \cdot X^k, \tag{4.7}$$

where

$$\begin{gathered} A = a_1 \cdot (\beta)^{a_2} \cdot (\text{Fr}_g)^{a_3} \cdot \left(\frac{\rho_g}{\rho_l}\right)^{a_4}, \\ B = a_5 \cdot (\beta)^{a_6} \cdot (\text{Fr}_g)^{a_7} \cdot \left(\frac{\rho_g}{\rho_l}\right)^{a_8}, \end{gathered} \tag{4.8}$$

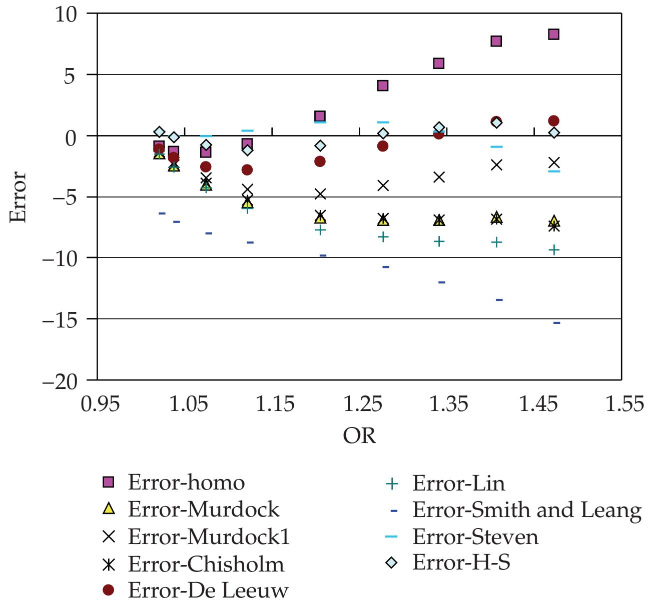


Figure 20: Comparison errors of models under $\beta = 0.60$, $P = 3$ MPa, and $\mathrm{Fr}_g = 1.5$.

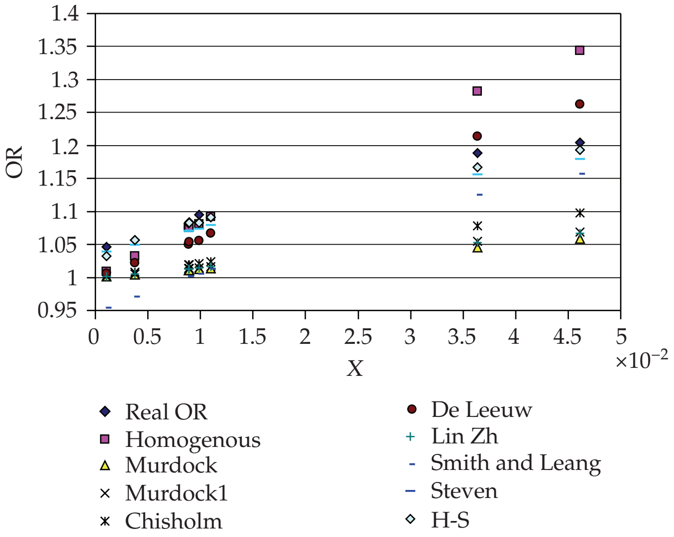


Figure 21: Comparison of models under $\beta = 0.70$, $P = 0.20$ MPa, and $\mathrm{Fr}_g = 1.7$.

where K is constant, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8 are undetermined coefficient, which will be determined through experimental data. The fit coefficient showed in Table 7.

Table 7 is the coefficient n fit by independent data from TJU low-pressure wet gas loop and National Engineering Laboratory high-pressure wet gas loop. Using exponent n and coefficient C_g for H-S over-reading model, over 98% data set will express the prediction error within ±5%, and the maximum error within ±6.5%. See Figures 11 and 12.

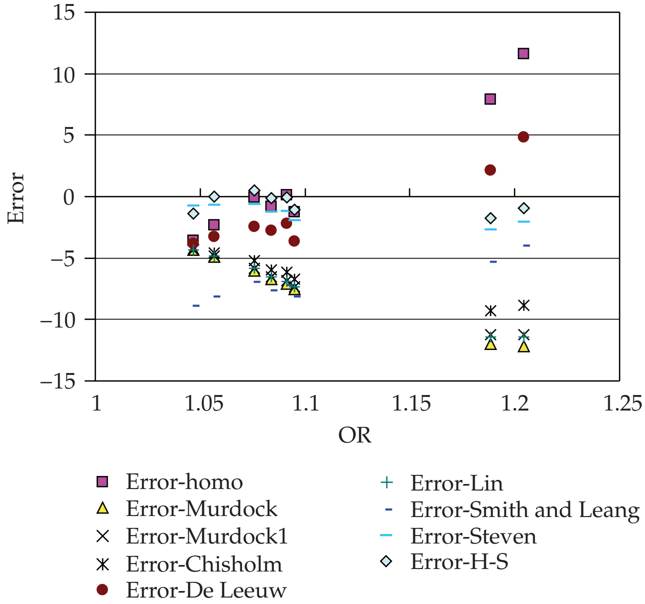


Figure 22: Comparison errors of models under β =0.70, P = 0.20 MPa, and Fr_g = 1.7.

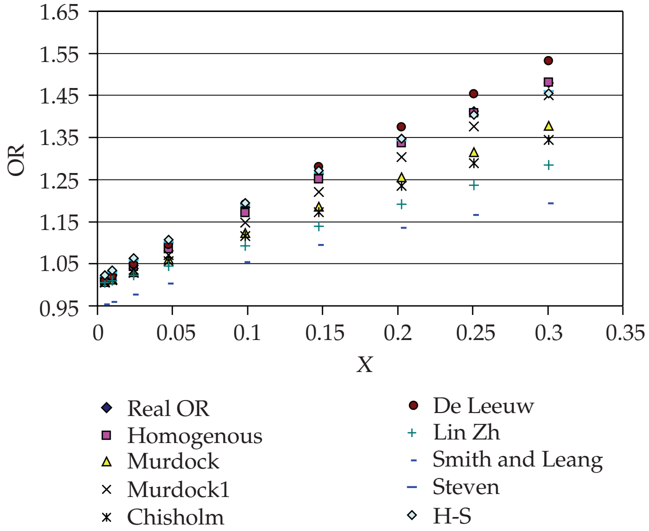


Figure 23: Comparison of models under β = 0.75, P = 6 MPa, and Fr_g = 3.5.

4.3. Comparison of H-S OR Model and the Eight Previous OR Models

Compare new model to 8 old models with the condition of pressure P varied from 0.15 to 6.0 MPa, beta ratio varied from 0.4 to 0.75, gas densimetric Froude number Fr_g varied from 0.5 to 5.5, the modified Lockhart-Martinelli parameter X varied from 0.002 to 0.3, the ratio of the gas to total mass flow rate x varied from 0.5 to 0.99. The data used for comparison is independent data different from training data. A Part of independent data was obtained from

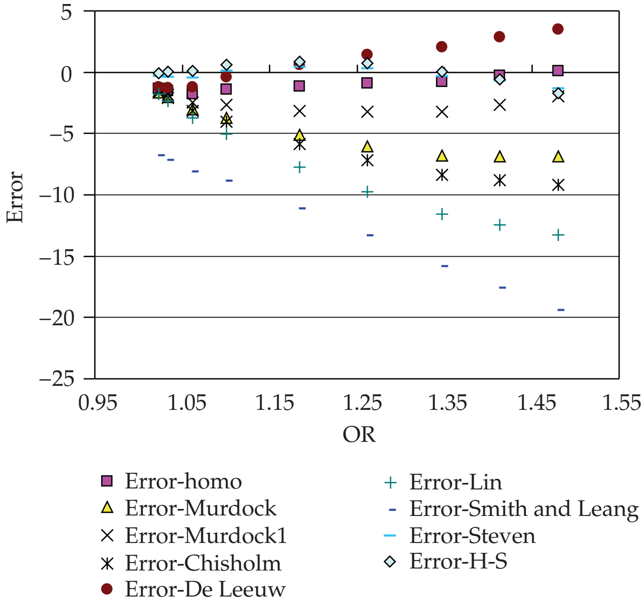


Figure 24: Comparison errors of models under $\beta = 0.75$, $P = 6\,\mathrm{MPa}$, and $\mathrm{Fr}_g = 3.5$.

NEL's report. Figures 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, and 24 are a part of the compared results.

These figures show that H-S model can accurately predict Venturi OR in all kinds of flow conditions, the error of H-S wet gas model is stable with OR increasing, and within 5%. It again proved that new wet gas model has good adaptability and wide application range. Particularly, as the wet gas flow fluctuate intensively under low pressure, all old OR models cannot predict OR accurately, the absolute of maximum error almost reached 40%. However, the new wet gas model reflects this change perfectly, the prediction OR has the same distribution with real OR. This is mainly because the homogenous model can well reflect the fluctuation of real OR, and the H-S model has inherited this ability. NEL' data have evidence trends because it is obtained in middle and high pressure. Even though, old correlations predicted errors are also large than H-S correlation, they varied from 10% to −35%.

5. Conclusions

Separation and homogeneous assumptions reflect the wet gas flow character, so a correlation combining these two assumptions performed well than each single one. The H-S model has inherited merits of homogeneous correlation and separation correlation, and can predict Venturi over reading accurately with the conditions of pressure varied from 0.15 to 6 MPa, beta ratio varied from 0.4 to 0.75, gas densimetric Froude number varied from 1 to 5.5, the modified Lockhart-Martinelli parameter varied from 0.002 to 0.3, the ratio of the gas to total mass flow rate varied from 0.5 to 0.99. The prediction error of H-S model is within ±6.5%.

Acknowledgments

This work was supported by the High-Tech Research and Development Program of China, numbers 2006AA04Z167, 2007AA04Z180, and supported in part by the National Natural

Science Foundation of China under the project Grant no. 60573125. The author would like to thank National Engineering Laboratory for providing the reports and professional advice on the web.

References

[1] J. Agar and D. Farchy, "Wet gas metering using dissimilar flow sensors: theory and field trial results," in *Proceedings of the SPE Annual Technical Conference and Exhibition*, pp. 1–6, San Antonio, Tex, USA, September-October 2002, SPE 77349.

[2] A. Lupeau, B. Platet, P. Gajan, A. Strzelecki, J. Escande, and J. P. Couput, "Influence of the presence of an upstream annular liquid film on the wet gas flow measured by a Venturi in a downward vertical configuration," *Flow Measurement and Instrumentation*, vol. 18, no. 1, pp. 1–11, 2007.

[3] Y. Geng, J. Zheng, and T. Shi, "Study on the metering characteristics of a slotted orifice for wet gas flow," *Flow Measurement and Instrumentation*, vol. 17, no. 2, pp. 123–128, 2006.

[4] J. W. Murdock, "Two-phase flow measurements with orifices," *Journal of Basic Engineering*, vol. 84, no. 4, pp. 419–433, 1962.

[5] R. V. Smith and J. T. Leang, "Evaluations of correlations for two-phase, flowmeters three current-one new," *Journal of Engineering for Power*, vol. 97, no. 4, pp. 589–593, 1975.

[6] D. Chisholm, "Research note: two-phase flow through sharp-edged orifices," *Journal of Mechanical Engineering Science*, vol. 19, no. 3, pp. 128–130, 1977.

[7] Z. H. Lin, "Two-phase flow measurements with sharp-edged orifices," *International Journal of Multiphase Flow*, vol. 8, no. 6, pp. 683–693, 1982.

[8] G. Toma, "Practical test-functions generated by computer algorithms," in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA '05)*, vol. 3482 of *Lecture Notes in Computer Science*, pp. 576–584, Singapore, May 2005.

[9] R. de Leeuw, "Wet gas flow measurement using a combination of Venturi meter and a tracer technique," in *Proceedings of the 12th North Sea Flow Measurement Workshop*, Peebles, Scotland, October 1994.

[10] R. N. Steven, "Wet gas metering with a horizontally mounted Venturi meter," *Flow Measurement and Instrumentation*, vol. 12, no. 5-6, pp. 361–372, 2001.

[11] NEL, "Effects of two-phase flow on single phase flow meters," *Flow Measurement Guidance Note*, no. 3, pp. 1–3, July 1997.

[12] NEL, *The Evaluation of Dry Gas Meters in Wet Gas Condition*, NEL, London, UK, 2002.

[13] NEL, *The Evaluation of Wet Gas Metering Technologies for Offshore Application: Part1-Differential Pressure Meters, Flow Measurement Guidance Note*, no. 40, NEL, London, UK, Feb 2003.

[14] C. Britton, W. Seidl, and J. Kinney, "Experimental, wet gas data for a Herschel style Venturi," in *Proceedings of the 5th International Symposium on Fluid Flow Measurement*, pp. 1–8, Arlington, Va, USA, April 2002.

[15] T. Kegel, "Wet gas measurement," in *Proceedings of the 4th CIATEQ Seminar on Advanced Flow Measurement*, pp. 1–7, Boca del Rio, Mexico, July 2003.

[16] F. Lide, Z. Tao, and J. Ningde, "A comparison of correlations used for Venturi wet gas metering in oil and gas industry," *Journal of Petroleum Science and Engineering*, vol. 57, no. 3-4, pp. 247–256, 2007.

[17] G. F. Hewitt, *Measurement of Two Phase Flow Parameter*, Academic Press, New York, NY, USA, 1978.

[18] G. B. Wallis, *One-Dimensional Two-Phase Flow*, McGraw-Hill, New York, NY, USA, 1969.

[19] C. Zhihang, C. Bolin, and Z. Zaisan, *Gas-Liquid Two Phase Flow and Heat Transfer*, Mechanical Industry Press, Peking, China, 1983.

[20] L. Zonghu, W. Shuzhong, and W. Dong, *Gas-Liquid Two Phase Flow and Bioling Heat Transfer*, Xi'An Jiao Tong University Press, Xi'An, China, 2003.

[21] L. Yi, *The study of wet gas measurement technology for applications: Venturi tube*, Master's Degree Dissertation, Tianjin University, Tianjin, China, 2005.

[22] ISO 5167-3:2003, "Measurement of fluid flow by means of pressure differential devices inserted in circular cross-section conduits running full—part 3: nozzles and Venturi nozzles".

[23] ISO 5167-1:2003, "Measurement of fluid flow by means of pressure differential devices inserted in circular cross-section conduits running full—part 1: general principles and requirements".

[24] ISO 5167-4:2003, "Measurement of fluid flow by means of pressure differential devices inserted in circular cross-section conduits running full—part 4: Venturi tubes".

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 543457, 15 pages
doi:10.1155/2008/543457

Research Article

Detection of Short-Step Pulses Using Practical Test-Functions and Resonance Aspects

Alexandru Toma and Cristian Morarescu

Corner Soft Technologies, 23 George Macarovici St., Bucharest 6, 060142, Romania

Correspondence should be addressed to Cristian Morarescu, c1mora2@yahoo.com

Received 16 May 2008; Accepted 3 June 2008

Recommended by Ming Li

An important aspect in modeling dynamic phenomena consists in measuring with higher accuracy some physical quantities corresponding to the dynamic system. Yet for measurements performed on limited time interval at high working frequency, certain intelligent methods should be added. The high working frequency requires that the measurement and data processing time interval should be greater than the time interval when the step input is received, so as to allow an accurate measurement. This paper will show that an intelligent processing method based on oscillating second-order systems working on limited time interval can differentiate between large step inputs (which are active on the whole limited time interval) and short step inputs (which are active on a time interval shorter than the limited working period). Some resonance aspects (appearing when the input frequency is close to the working frequency of the oscillating second-order system) will be also presented.

1. Introduction

Filtering and sampling devices usually consist of asymptotically stable systems, sometimes an integration of the output over a certain time interval being added. Yet such structures are very sensitive at random variations of the integration period, being recommended for the signal which is integrated to be approximately equal to zero at the end of the integration period. For this reason, oscillating systems for filtering the received signal should be used, so as the filtered signal and its slope to be approximately zero at the end of a certain time interval (at the end of an oscillation). For avoiding instability of such oscillating systems on extended time intervals, certain electronic devices (gates) controlled by computer commands should be added, so as to restore the initial null conditions for the oscillating system before a new working cycle to start [1].

The filtering performances of asymptotically stable systems are determined by their transfer function. A filtering and sampling device consisting of low-pass filters of first or second order having the transfer function

$$H(s) = \frac{1}{T_0 s + 1} \tag{1.1}$$

(for a first-order system) and

$$H(s) = \frac{1}{T_0^2 s^2 + 2bT_0 s + 1} \tag{1.2}$$

(for a second-order system) attenuates an alternating signal of angular frequency $\omega \gg \omega_0 = 1/T_0$ about ω/ω_0 times (for a first-order system) or about $(\omega/\omega_0)^2$ times (for a second-order system). The response time of such systems at a continuous useful signal is about $4-6T_0$ ($5T_0$ for the first-order system and $4T_0/b$ for the second-order system). If the signal given by the first- or second-order system is integrated over such a period, a supplementary attenuation for the alternating signal of about $4-6\omega/\omega_0$ can be obtained.

But such structures are very sensitive at the random variations of the integration period (for unity-step input, the signal which is integrated is equal to unity at the sampling moment of time), and the use of oscillators with a very high accuracy cannot solve the problem due to switching phenomena appearing at the end of the integration period (when an electric current charging a capacitor is interrupted).

These random variations cannot be avoided if we use asymptotically stable filters. By the other hand, an improvement in an electrical scheme used for integrators in analog signal processing (see [2, 3]) cannot lead to a significant increasing in accuracy, as long as such electronic devices perform the same task (the system has the same transfer function). There are also known techniques for reducing the switching noise in digital systems, but such procedures can be applied only after the analog signal is filtered and sampled, so as to be prepared for further processing. So we must give attention to some other kind of transfer functions and to analyze their properties in case of filtering and sampling procedures, similar to wavelets analysis presented in [4, 5].

Mathematically, an ideal solution consists in using an extended Dirac function for multiplying the received signal before the integration (see [1]), but is very hard to generate such extended Dirac functions (a kind of acausal pulses) using nonlinear differential equations for (i) symmetrical pulses (see [6]) or (ii) asymmetrical pulses (see [7] for more details).

A heuristic algorithm for generating practical test functions using MATLAB procedures was presented in [6]. First it has been shown that ideal test functions cannot be generated by differential equations, being emphasized the fact that differential equations can only generate functions similar to test functions (defined as practical test functions). Then a step-by-step algorithm for designing the most simple differential equation able to generate a practical test function was presented, based on the invariance properties of the differential equation and on standard MATLAB procedures. The result of the algorithm consists in a system working at the stability limit from initial null conditions, on limited time intervals, the external signal representing the free term in the differential equation corresponding to the input of the oscillating system. Such a system could be built using standard components and operational amplifiers. However, the previously mentioned study [6] did not investigate the behavior of such an oscillating system for an input represented by a short-step pulse. These aspects will be studied in this paper. Finally, supplementary resonance aspects (appearing when the input frequency is close to the working frequency of the oscillating second-order system) will be also presented.

2. Modeling transitions by practical test-functions integral aspects

While this study is based on robust integral procedures of practical test-functions for certain time intervals, certain basic integral aspects of practical test functions should be mentioned. These aspects are useful for modeling smooth transitions from a certain function of time to another on a limited time interval [6].

From basic mathematics, it is known that the product $\varphi(t)g(t)$ between a function $g(t)$ which belongs to C^{∞} class and a test-function $\varphi(t)$ which differs to zero on (a,b) is also a test-function because

(a) it differs to zero only on the time interval (a,b) where $\varphi(t)$ differs to zero (if $\varphi(t)$ is null, then the product $\varphi(t)g(t)$ is also null);

(b) the function $\varphi(t)g(t)$ belongs to the C^{∞} class of functions, while a derivative of a certain order k can be written as

$$\left(\varphi(t)g(t)\right)^{(k)} = \sum_{p=0}^{k} C_k^p \varphi(t)^{(p)} g(t)^{(k-p)} \tag{2.1}$$

(a sum of terms represented by a product of two continuous functions).

Yet for practical cases (when phenomena must be represented by differential equations), the $\varphi(t)$ test functions must be replaced by a practical test functions $f(t) \in C^n$ on R (for a finite n-considered from now on as representing *the order* of the practical test function) having the following properties:

(a) f is nonzero on (a), (b),

(b) f satisfies the boundary conditions $f^{(k)}(a) = f^{(k)}(b) = 0$ for $k = 0, 1, \dots, n$, and

(c) f restricted to (a,b) is the solution of an initial value problem (i.e., an ordinary differential equation on (a,b) with initial conditions given at some point in this interval).

The generation of such practical test functions is based on the study of differential equations satisfied by these test functions, with the initial moment of time chosen at a time moment close to the $t = a$ moment of time (when the function begins to present nonzero values).

By using these properties of practical test-functions, we obtain the following important result for a product $f(t)g(t)$ between a function $g(t)$ which belongs to C^{∞} class and a practical test-function of n order $f(t)$ which differs to zero on (a,b):

General property for product

The product $g(t)f(t)$ between a function $g(t) \in C^{\infty}$ and a practical test-function f of order n is represented by a practical test function of order n.

This is a consequence of the following two properties:

(a) the product $g(t)f(t)$ differs to zero only on the time interval (a,b) on which $f(t)$ differs to zero;

(b) the derivative of order k for the product $g(t)f(t)$ is represented by the sum

$$(f(t)g(t))^{(k)} = \sum_{p=0}^{k} C_k^p f(t)^{(p)} g(t)^{(k-p)} \tag{2.2}$$

which is a sum of terms representing products of two continuous functions for any $k \leq n$, (n being the order of the practical test-function f)—only for $k > n$ discontinuous functions can appear in the previous sum.

Now we will begin to study the integral properties of practical test functions of certain order. For this, we note that the integral $\varphi(t)$ of a test function $\phi(t)$ (which differs to zero on (a, b) interval) is a constant function on the time intervals $(-\infty, a]$ and $[b, +\infty)$; it presents a certain variation on the (a, b) time interval, from a constant null value to a certain Δ quantity corresponding to the final constant value. Moreover, all derivatives of order $k \leq n + 1$ for the integral function $F(t)$ are equal to zero for $t = a$ and $t = b$ (this can be easily checked by taking into account that all derivatives of order p for $f(t)$ are equal to zero at these time moments, for $p \leq n$, and a derivative of order p for $f(t)$ corresponds to a derivative of order $p+1$ for function $F(t)$, the integral function of $f(t)$). This suggests the possibility of using such integral functions for modeling smooth transitions from a certain state to another in different kind of applications, when almost all derivatives of a certain function are equal to zero at the initial moment of time.

For modeling such a transition, we analyze the general case when a function f and a finite number of its derivatives $f^{(1)}, f^{(2)}, \dots f^{(n)}$ present variations from null values to values $\Delta, \Delta_1, \Delta_2, \dots \Delta_n$ on the time interval $[-1, 1]$. We begin by looking for a function f_n which should be added to the null initial function so as to obtain a variation Δ_n for the derivative of n order.

By multiplying the bump-like function

$$\varphi(\tau) = \begin{cases} C \exp\left(\dfrac{1}{\tau^2 - 1}\right), & \text{if } \tau \in (-1, 1), \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

(a test-function on $[-1, 1]$) with the variation Δ_n of the derivative of n order and by integrating this product $n + 1$ times we obtain

(i) after the first integration: a constant value equal to Δ_n at the time moment $t = 1$ (while the integral of the bump-like test function on $[-1, 1]$ is equal to 1), and a null variation on $(1, +\infty)$;

(ii) after the second integration (when we integrate the function obtained at previous step): a term equal to $\Delta_n(t - 1)$ and a term equal to a constant value c_{11} (a constant of integration) on the time interval $(1, +\infty)$;

(iii) after the $n + 1$ integration: a term equal to $\Delta_n(t - 1)^n/n!$ and a sum of terms having the form $c_{1i}(t - 1)^i/i!$ for $i \in N,\ i < n$ (c_{ni} being constants of integration) on the time interval $(1, +\infty)$.

All previous constants of integration are determined by integrating the test function on $[-1, 1]$. The procedure continues by looking for the other functions $f_{n-1}, f_{n-2} \dots$ which must be added to the initial null function. However, we must take care to the fact that the function f_n previously obtained has nonzero variations $d_{n-1}, d_{n-2}, \ldots d_1$ for its derivatives of order $n-1, n-2, \ldots 1$ on the working interval and so we must subtract these values from the set $\Delta_{n-1}, \Delta_{n-2}, \ldots \Delta_1$ before passing to the next step.

Then we multiply the bump-like function with the corrected value

$$\Delta'_{n-1} = \Delta_{n-1} - d_{n-1} \tag{2.4}$$

and by integrating this product n times we obtain in a similar manner a function with a term equal to $\Delta'_{n-1}(t-1)^{n-1}/(n-1)!$ and a sum of terms having the form $c_{2i}(t-1)^i/i!$ for $i \in N, i < n-1$ (c_{ni} being constants of integration) on the time interval $(1, +\infty)$. It can be noticed that the result obtained after n integration possess the $n-1$ order derivative equal to Δ'_{n-1}, a smooth transition for this derivative from the initial null value being performed. So the second function which must be added to the initial null function is the integral of $n-1$ order for the bump-like function multiplied by this variation $\Delta_{n-1} - d_{n-1}$ (the function being noted as f_{n-1}). This function f_{n-1} has a null value for the derivative of n order for $t > 1$, so the result obtained at first step is not affected. We must take care again to the fact that the function f_{n-1} previously obtained has nonzero variations $d^1_{n-1}, d^1_{n-2}, \ldots d^1_1$ for its derivatives of order $n-1, n-2, \ldots 1$ and so we must once again subtract these values from the previously corrected set $\Delta_{n-1}-d_{n-1}, \Delta_{n-2}-d_{n-2}, \ldots \Delta_1-d_1$ before passing to the next step. Finally we obtain all functions $f_{n+1}, f_n, \ldots f_1$ which represent the terms of function f modeling the smooth transition from an initial null function to a function having a certain set of variations for a finite number of its derivatives on a small time interval. The procedure can be also applied for functions possessing a finite number of derivatives within a certain time interval by time reversal (t being replaced with $-t$). More details regarding possible applications of such a procedure can be found in [6].

3. The oscillating second-order system for the case of short-step inputs

After presenting basic aspects regarding integral properties of practical test-functions, we will analyze the behavior of a system able to generate a practical test-function for signal processing when its command is represented by a short-step pulse. Unlike aspects presented in previous paragraph, the step change appears for the command function $u(t)$, and the dynamical behavior on a limited time interval should be performed. We are searching for a robust integral procedure (with null values of the function which is integrated at the beginning and at the end of the interval of integration) so as the sampled values to be further processed for determining the amplitude and the time length of the short-step pulse.

For a robust filtering and sampling procedure based on an integration on a limited time interval, a search for a system having the following property was performed in a rigorous manner in [8]: starting to work from initial null conditions, for a unity step input it must generate an output and a derivative of this output equal to zero at a certain moment of time (the condition for the derivative of the output to be equal to zero has been added so as the slope and the first derivative of the slope for the signal which is integrated to be equal to zero at the sampling moment of time, when the integration is interrupted). It was finally shown that the simplest structure possessing such properties is represented by an oscillating second-order system having the transfer function

$$H_{\text{osc}} = \frac{1}{T_0^2 s^2 + 1} \tag{3.1}$$

receiving a step input and working on the time interval $[0, 2\pi T_0]$. For initial conditions equal to zero, the response of the oscillating system at a step input with amplitude A will have the form

$$y(t) = A\left(1 - \cos\left(\frac{t}{T_0}\right)\right). \tag{3.2}$$

By integrating this result on the time interval $[0, 2\pi T_0]$, we obtain the result $2\pi A T_0$, and we can also notice that the quantity which is integrated and its slope are equal to zero at the end of the integration period. Thus the influence of the random variations of the integration period (generated by the switching phenomena) is practically rejected.

This oscillating system attenuates about $(\omega/\omega_0)^2$ times such an input, and the influence of the integrator consists in a supplementary attenuation of about

$$\left[(1/(2\pi))(\omega/\omega_0)\right] \tag{3.3}$$

times. The oscillations having the form

$$y_{\text{osc}} = a \sin(\omega_0 t) + b\cos(\omega_0 t) \tag{3.4}$$

generated by the input alternating component have a lower amplitude and give a null result after an integration over the time interval $[0, 2\pi T_0]$.

These results have shown that such a structure provides practically the same performances as a structure consisting of an asymptotically stable second-order system and an integrator (response time of about $6T_0$, an attenuation of about $(1/6)(\omega/\omega_0)^3$ times for an alternating component having frequency ω), moreover being less sensitive at the random variations of the integration period. For restoring the initial null conditions after the sampling procedure (at the end of the working period), some electronic devices must be added. Yet the previous analysis is valid for extended step inputs, which are active on the whole working interval (the integration period).

We will continue the analysis of this structure by considering that the input is represented by a unity short-step pulse (instead of a unity step-pulse) which differs to zero on the time interval $[0, \tau]$. This means that the input u can be represented under the following form:

$$\begin{aligned} u(t) &= 1, \quad \text{for } t \in [0, \tau], \\ u(t) &= 0, \quad \text{for } t > \tau, \end{aligned} \tag{3.5}$$

or, using the Heaviside function

$$u(t) = h(\tau - t) \quad \text{for } t \in [0, \infty), \tag{3.6}$$

where $h(\tau)$ corresponds to the function $1/s$ if we apply the Laplace transformation.

The transfer function of the second-order oscillating system is

$$H(s) = \frac{1}{T_0^2 s^2 + 1}. \tag{3.7}$$

On the time interval $[0, \tau]$, the output of the second-order oscillating system is represented (using the Laplace transformation) as

$$y(s) = H(s)u(s) = \frac{1}{T_0^2 s^2 + 1}\frac{1}{s} \tag{3.8}$$

which corresponds to the output

$$y(t) = \left(1 - \cos\left(\frac{t}{T_0}\right)\right) \tag{3.9}$$

which can be written as

$$y(t) = 1 - \cos(\omega_0 t), \tag{3.10}$$

where $\omega_0 = 2\pi/T_0$. When the action of the external unity pulse ceases (for $t = \tau$) the output $y(t)$ is

$$y(\tau) = 1 - \cos(\omega_0 \tau) \tag{3.11}$$

and the derivative of $y(t)$ is

$$y'(\tau) = \omega_0 \sin(\omega_0 \tau). \tag{3.12}$$

These values, $y(\tau)$ and $y'(\tau)$, represent the initial values for the free oscillations of the second-order oscillating system generated for $t > \tau$ (when the input command $u(t) = 0$). These free oscillations have the angular velocity ω_0, and thus the output $y(t)$ for $t > \tau$ will have the form

$$y(t) = C \sin(\omega_0 t + \phi), \quad \text{for } t > \tau. \tag{3.13}$$

The quantities C and ϕ (amplitude and initial phase of free oscillations) should be determined using the initial conditions for $t = \tau$:

$$y(\tau) = 1 - \cos(\omega_0 \tau), \qquad y'(\tau) = \omega_0 \sin(\omega_0 \tau). \tag{3.14}$$

This implies

$$1 - \cos(\omega_0 \tau) = C \sin(\omega_0 \tau + \phi), \qquad \omega_0 \sin(\omega_0 \tau) = C\omega_0 \cos(\omega_0 \tau + \phi). \tag{3.15}$$

By dividing the second equality with ω_0, squaring both equalities and summing left-hand sides and right-hand sides of both squared equalities we obtain

$$[1 - \cos(\omega_0 \tau)]^2 + \sin^2(\omega_0 \tau) = C^2 \tag{3.16}$$

and (while $\sin^2(\omega_0 \tau) + \cos^2(\omega_0 \tau) = 1$)

$$2 - 2\cos(\omega_0 \tau) = C^2. \tag{3.17}$$

While

$$\cos(\omega_0\tau) = 1 - 2\sin^2\left(\frac{\omega_0\tau}{2}\right) \tag{3.18}$$

by substituting $\cos(\omega_0\tau)$ with the right-hand side of the above equality, the result is

$$4\sin^2\left(\frac{\omega_0\tau}{2}\right) = C^2. \tag{3.19}$$

It results

$$C = 2\sin\left(\frac{\omega_0\tau}{2}\right). \tag{3.20}$$

(C is a positive quantity, because $\omega_0\tau \in [0, 2\pi] \rightarrow \omega_0\tau/2 \in [0, \pi] \rightarrow \sin(\omega_0\tau/2) \geq 0$). The phase ϕ can be obtained using the two equations determined by the initial conditions:

$$1 - \cos(\omega_0\tau) = C\sin(\omega_0\tau + \phi), \qquad \sin(\omega_0\tau) = C\cos(\omega_0\tau + \phi) \tag{3.21}$$

(the second equation resulting by dividing previous equation of $y'(\tau)$ to ω_0).

By dividing left-hand side of first equality to left-hand side of second equality, and right-hand side of first equality to right-hand side of second equality, it results

$$\frac{1 - \cos(\omega_0\tau)}{\sin(\omega_0\tau)} = \tan(\omega_0\tau + \phi). \tag{3.22}$$

Using equalities

$$\cos(\omega_0\tau) = 1 - 2\sin^2\left(\frac{\omega_0\tau}{2}\right), \qquad \sin(\omega_0\tau) = 2\sin\left(\frac{\omega_0\tau}{2}\right)\cos\left(\frac{\omega_0\tau}{2}\right), \tag{3.23}$$

it results by substituting $\cos(\omega_0\tau)$, $\cos(\omega_0\tau)$ with the above expressions

$$\tan\left(\frac{\omega_0\tau}{2}\right) = \tan(\omega_0\tau + \phi). \tag{3.24}$$

So ϕ is obtained as

$$\phi = -\left(\frac{\omega_0\tau}{2}\right) \tag{3.25}$$

and the output $y(t)$ corresponding to the free oscillations of the system for $t > \tau$ (when the action of the external short-step command u has ceased) can be written as

$$y(t) = 2\sin\left(\frac{\omega_0\tau}{2}\right)\sin\left(\omega_0 t - \frac{\omega_0\tau}{2}\right). \tag{3.26}$$

4. Algorithm for detecting short-step pulses

While the signal processing system is linear, in case of a short-step pulse of amplitude A defined on time interval $[0, \tau]$ the output $y(t)$ of the system will be multiplied by A as related to the output obtained in case of a unity short-step input (presented in previous paragraph). Thus, the output $y(t)$ will be

$$\begin{aligned} & A(1 - \cos(\omega_0 t)), \quad \text{for } t \in [0.\tau], \\ y(t) = {} & 2A \sin\left(\frac{\omega_0 \tau}{2}\right) \sin\left(\omega_0 t - \frac{\omega_0 \tau}{2}\right), \quad \text{for } t > \tau. \end{aligned} \tag{4.1}$$

This output, $y(t)$, is equal to zero at two time moments $t1$ and $t2$ after time moment τ. At time moment $t1$,

$$\left(\omega_0 t1 - \frac{\omega_0 \tau}{2}\right) = \pi; \tag{4.2}$$

and at time moment $t2$,

$$\left(\omega_0 t2 - \frac{\omega_0 \tau}{2}\right) = 2\pi. \tag{4.3}$$

These imply that

$$\begin{aligned} t1 &= \frac{\tau}{2} + \frac{\pi}{\omega_0}, \\ t2 &= \frac{\tau}{2} + 2\left(\frac{\pi}{\omega_0}\right). \end{aligned} \tag{4.4}$$

We must check whether both $t1, t2$ are greater than τ. First we check the inequality $t1 = \tau/2 + \pi/\omega_0 > \tau$. This is equivalent to $\tau/2 < \pi/\omega_0$, $\tau < 2\pi/\omega_0 = T_0$, where T_0 represents the period of the second-order oscillating system. It is obvious that $\tau < T_0$, while we have considered that the short-step pulse has nonzero values for $t \in (0, T_0)$. Thus $t1 > \tau$, and while $t2 > t1$ it results that $t2 > t1 > \tau$.

The signal processing system will perform the integration of output $y(t)$ on two different time intervals. The first value $I1$ is obtained by an integration of $y(t)$ on the time interval $[0, t2]$. It results

$$\begin{aligned} I1 &= \int_0^{t2} y(t)dt = \int_0^{\tau} A(1 - \cos(\omega_0 t))dt + \int_{\tau}^{t2} 2A \sin\left(\frac{\omega_0 \tau}{2}\right) \sin\left(\omega_0 t - \frac{\omega_0 \tau}{2}\right)dt, \\ I1 &= A\left(\tau - \frac{1}{\omega_0} \sin(\omega_0 \tau) + 2A \frac{1}{\omega_0} \sin\left(\frac{\omega_0 \tau}{2}\right)\left(-\cos(2\pi) + \cos\left(\frac{\omega_0 \tau}{2}\right)\right)\right), \end{aligned} \tag{4.5}$$

(while $\omega_0 t2 = 2\pi$). Then we obtain

$$I1 = A\tau - A\frac{1}{\omega_0} \sin(\omega_0 \tau) + A\frac{2}{\omega_0} \sin\left(\frac{\omega_0 \tau}{2}\right)\left(\cos\left(\frac{\omega_0 \tau}{2}\right) - 1\right) \tag{4.6}$$

which can be finally written as

$$I1 = A\tau - A\frac{2}{\omega_0}\sin\left(\frac{\omega_0\tau}{2}\right). \tag{4.7}$$

The second integral $I2$ is performed by integrating $y(t)$ on the time interval $[t1, t2]$. On this time interval, the output $y(t)$ is represented by a free oscillation

$$y(t) = 2A\sin\left(\frac{\omega_0\tau}{2}\right)\sin\left(\omega_0 t - \frac{\omega_0\tau}{2}\right), \tag{4.8}$$

so $I2$ is determined by

$$I2 = \int_{t1}^{t2} 2A\sin\left(\frac{\omega_0\tau}{2}\right)\sin\left(\omega_0 t - \frac{\omega_0\tau}{2}\right)dt; \tag{4.9}$$

and taking into account that

$$\begin{aligned} \omega_0 t - \frac{\omega_0\tau}{2} &= \pi, \quad \text{for } t = t1, \\ \omega_0 t - \frac{\omega_0\tau}{2} &= 2\pi, \quad \text{for } t = t2, \end{aligned} \tag{4.10}$$

it results $I2$ as

$$I2 = -\frac{4A}{\omega_0}\sin\left(\frac{\omega_0\tau}{2}\right). \tag{4.11}$$

The values $I1, I2$ allow us to determine A, τ by robust integrations (the values of the function $y(t)$ which is integrated are zero at the beginning and the end of the time interval used for integration). A quantity $I0$ can be determined as

$$I0 = I1 - \frac{I2}{2} = A\tau; \tag{4.12}$$

and a quantity R can be determined as

$$R = -\frac{I2}{2I0} = \frac{(4A/\omega_0)\sin(\omega_0\tau/2)}{2A\tau} \tag{4.13}$$

which can be written as

$$R = \frac{\sin(\omega_0\tau/2)}{\omega_0\tau/2} = \operatorname{sinc}\left(\frac{\omega_0\tau}{2}\right). \tag{4.14}$$

This means that after performing the robust integration of $y(t)$ in order to obtain the sampled values $I1, I2$ we can compute $I0 = I1 - I2/2$ and then $I0 = -I2/(2I0)$. While

$$I0 = \operatorname{sinc}\left(\frac{\omega_0\tau}{2}\right), \tag{4.15}$$

we can determine the time interval τ of the received short-step pulse using $I0$ and a mathematical memory (the quantity ω_0 being known).

Next quantity A (the amplitude of this received short-step pulse) can be determined using τ (determined at previous computation) and $I0 = A\tau$.

```
At $t = 0$ Start Integration for $I1$
if output = 0
then
Start Integration for $I2$
if output = 0
Stop Integration for $I1$ and Integration for $I2$
Determine $A$ and $\tau$ using sampled values for $I1$ and $I2$
```

Algorithm 1

Taking into account the previous considerations, the algorithm for detecting short-step pulses consists in the steps shown in Algorithm 1.

5. The problem of initial conditions and resonance aspects

In previous paragraphs has been analyzed the case when a certain short-step pulse is received at the beginning of a working interval—this pulse presenting nonzero values on a limited time interval $(0, \tau)$. Yet in applications, the signal processing system can receive a sequence of step pulses with different time lengths. Due to this reason, the whole procedure must be adapted for the case when a certain step input with amplitude D is received on a certain time interval $(0, \sigma)$ (the first part of the working interval) and a different step input with amplitude B appears on the time interval $[\sigma, T_0]$ (the second part of the time interval).

Under these circumstances the output $y(t)$ could be represented by

$$\begin{aligned} y(t) &= D\bigl(1 - \cos(\omega_0 t)\bigr), \quad \text{for } t < \sigma, \\ y(t) &= B\bigl(1 - \cos(\omega_0(t - \sigma))\bigr) + y_{\text{osc}}, \quad \text{for } t \geq \sigma, \end{aligned} \tag{5.1}$$

where

$$y_{\text{osc}} = 2D \sin\left(\frac{\omega_0 \sigma}{2}\right) \sin\left(\omega_0 t - \frac{\omega_0 \sigma}{2}\right) \tag{5.2}$$

represents the free oscillations of the system generated by the short-step pulse with amplitude D (which has ceased its action at time moment σ). It can be noticed that it is very difficult to analyze these outputs by performing certain integral operations in order to determine the parameters D, B, σ. However, the whole procedure can be simplified in a significant manner if we observe that the amplitude D of the first step input can be predicted by the signal processing system (by considering that the pulse with amplitude D has been received by the system on a previous working interval and its action continues at the beginning of the analyzed working interval). This suggests the possibility of adjusting the input command for the second-order oscillating system by subtracting quantity D from $u(t)$ and thus the output of the oscillating system will become

$$\begin{aligned} y(t) &= 0, \quad \text{for } t < \sigma, \\ y(t) &= (B - D)\bigl(1 - \cos(\omega_0(t - \sigma))\bigr), \quad \text{for } t \geq \sigma. \end{aligned} \tag{5.3}$$

The quantity y_{osc} vanishes while the oscillating second-order system has initial null conditions and a null command for $t < \sigma$.

At first sight it looks like the system receives a nonzero command at the end of the working time interval and the results presented in previous paragraph cannot be applied. Yet we can observe that an integration performed on the first time interval $[0, \sigma]$ generates a null result. Due to this reason, we have to adjust the integration procedure by

(i) starting the integration for $I1$ and $I2$ at the beginning of a new working cycle, with the input command considered as $w(t) = u(t) - A$ (the previous value for the amplitude of the received pulse is subtracted from received pulse $u(t)$);

(ii) if a step variation of amplitude for the input command $w(t)$ is detected then

  (a) *stop* the action of adjusted input command $w(t) = u(t) - A$ upon the oscillating system at time moment T_0 (the output $y(t)$ of the system will be represented by free oscillations),

  (b) *continue* the integration for $I1$ and $I2$ after the time moment T_0 (which would have been the end of the working cycle in case that the step variation of input has been not detected);

(iii) stop the integration procedure for $I1$ when $y(t)$ first time equals zero and the integration procedure for $I2$ when $y(t)$ second time equals zero.

Note that the time moments when the integration procedures cease are not affected by noise, while the output $y(t)$ is represented just by free oscillations of the second-order system (after the time moment T_0 the action of $u(t)$ upon the system ceases);

(iv) Determine σ using

$$\begin{gathered} I0 = -I2/(2I0), \quad I0 = \operatorname{sinc}\left(\frac{\omega_0 \tau}{2}\right), \quad T_0 - \sigma = \tau, \\ I0 = A\tau, \quad B - D = A, \quad B = D + A. \end{gathered} \tag{5.4}$$

Note: by translating the time origin from the beginning of a working interval to the time moment when the step change $(B - D)$ appears, we can consider that a short-step pulse with amplitude $(B - D)$ acts upon the second-order system at time moment zero, from null initial conditions, on the time interval $(0, T_0 - \sigma)$—thus the quantities τ and A corresponding, respectively, to the time length of the received short-step pulse and to its amplitude in previous paragraph should be replaced by $T_0 - \sigma$ and $(B - D)$.

All presented aspects are valid if the system receives a step pulse (which can be represented by an extended step pulse as presented in [1]) or by a short-step pulse (as presented in this study, in previous paragraphs). Filtering properties of the second-order oscillating system were studied in [1]. However, we must study also resonance aspects. The second-order system being an oscillating system, resonance aspects (appearing when the input $u(t)$ is represented by an alternating function $A\sin(\omega t + \varphi)$ with $\omega \approx \omega_0$) are very important. Instead of damped proper oscillations of angular frequency ω_0 with zero limit (as for an asymptotic stable second-order system), some proper oscillations of angular frequency ω_0 and with constant amplitude can be noticed as term in $y(t)$. These are added to the oscillations with angular frequency ω generated by the command function $u(t)$, both having a higher amplitude inversely proportional to $\omega^2 - \omega_0^2$. If $\omega = \omega_0$ (the limit case), for a command function

$$u(t) = A\sin\left(\omega_0 t + \varphi_0\right), \tag{5.5}$$

the output $y(t)$ is represented by

$$y(t) = E_1 t \sin\left(\omega_0 t + \varphi_1\right) + E_2 \sin\left(\omega_0 t + \varphi_2\right). \tag{5.6}$$

This function is hard to be processed by a signal processing system, working on a limited time interval. Moreover, in applications it is possible for the input $u(t)$ to be represented by a sum of alternating functions with different angular frequencies $\omega_i \approx \omega_0$. However, a signal processing procedure can be established for the case when $u(t)$ is represented by a sum of alternating functions with angular frequencies $\omega_i \approx \omega_0$ by taking into account the fact that for an input $u(t)$ represented by an alternating function $A\sin(\omega t + \varphi)$ the amplitude E of oscillations with angular frequency ω generated by this command function $u(t)$ is determined by

$$E(\omega) = \frac{A\omega_0^2}{\omega^2 - \omega_0^2} \tag{5.7}$$

which is a very sharp function. For this reason we can consider that in the case when the input $u(t)$ is represented by a sum of oscillations with different angular frequencies $\omega_i \approx \omega_0$, the output $y(t)$ of the second-order system will be represented by an oscillation with the angular frequency ω_j closest to ω_0 (generated by the received oscillation with angular frequency ω_j) and a proper oscillation of the second-order system (with angular frequency ω_0). Thus $y(t)$ could be represented by a sum of two oscillations.

The result of an integration of this output $y(t)$ on the working interval $[0, 2\pi T_0]$ would depend just on the oscillation with angular frequency ω_j, while the oscillation with angular frequency ω_0 (with time period T_0) gives a null result by an integration on this period. The result of this integration could be used for determining certain parameters for the received oscillation with angular frequency ω_j. However, such an integration is not a robust integration, while the signal which is integrated is not equal to zero at the end of this working interval. A possible solution of this problem would consist in disconnecting the input signal after a certain time interval, so as to analyze (using robust integrations) the free oscillations of the second-order system after this moment (as presented in previous paragraph). A faster procedure could consist in a previous adjustment of initial conditions, so as the free oscillations not to appear.

Theoretically, this can be done by using a set of identical oscillating second-order systems (receiving the same input command $u(t)$) with initial conditions adjusted to different values. The system generating a single oscillation with angular frequency ω_j would be selected by checking the following condition:

$$\frac{d^2y}{dt^2} + \omega_j^2 y = 0. \tag{5.8}$$

However, the adjustment of two initial conditions at different values for a second-order system requires a large number of identical oscillators. Due to this reason, this method is inconvenient. A better choice would be represented by a delay systems of first order, with transfer function

$$H(s) = \frac{s}{s\sin\varphi + \omega_j \cos\varphi} \tag{5.9}$$

would transform a received oscillation $\sin(\omega_j t + \varphi)$ according to

$$V(s) = H(s)U(s) = \frac{s}{s\sin\varphi + \omega_j\cos\varphi}\frac{s\sin\varphi + \omega_j\cos\varphi}{s^2 + \omega_j^2} = \frac{s}{s^2 + \omega_j^2} \tag{5.10}$$

which corresponds to an output $v(t) = \cos(\omega_0 t)$. If this function $v(t)$ represents the command for the second-order oscillating system, the output $y(t)$ (for initial null conditions) will be

$$y(t) = (1/2)\omega_j t \sin\left(\omega_j t\right) \tag{5.11}$$

when ω_j is very close to ω_0 and can be approximated by this quantity. The function $y(t)$ presented in previous equation is suitable for a sequence of robust integration procedures on half-period time intervals:

$$\left[0, \pi/\omega_j\right], \left[\pi/\omega_j, 2\pi/\omega_j\right], \text{ etc.} \tag{5.12}$$

(it presents null values at the beginning and at the end of these intervals integration). The results of these procedures are proportional to the amplitude A of received oscillation with angular frequency ω_j (as can be easily noticed).

Unlike the possible solution based on adjustment of two initial conditions, this procedure requires a set of signal processing systems composed of different time-delaying systems adjusted according to a single parameter (the phase ϕ, while ω_j is supposed to be known) and identical second-order oscillating systems starting to work from initial null conditions. The number of required systems is less than in previous case, while a single parameter has to be adjusted at different values (the quantity φ in the time-delay systems). The corresponding output is selected by checking whether the results of these procedures (considered as positive quantities) vary according to a linear mathematical law (as required by the integration of $y(t) = (1/2)\omega_j t \sin(\omega_j t)$ on a sequence of half-period time intervals).

One major disadvantage of this method has to be mentioned: the function $y(t) = (1/2)\omega_j t \sin(\omega_j t)$ equals zero at the beginning of signal processing time interval ($t = 0$) and presents a small slope. So an extended time interval is necessary for obtaining significant results using robust integration procedures. If we need a fast signal processing, we can simply use a set of identical oscillating second-order systems, with different initial conditions. The greatest amplitude for the output oscillation (considered as a sum of an oscillation with angular frequency ω_j and an oscillation with angular frequency ω_0, $\omega_j \approx \omega_0$) corresponds to the case when the two oscillations are in-phase. By detecting the output presenting in-phase oscillations, we can establish the amplitude and phase for the received signal using the initial conditions for the second-order system generating this output.

6. Conclusions

An important aspect in modeling dynamic phenomena consists in measuring with higher accuracy some physical quantities corresponding to the dynamic system. Yet for measurements performed on limited time interval at high working frequency, certain intelligent methods should be added. The high working frequency requires that the measurement and data processing time interval should be greater than the time interval when the step input is received, so as to allow an accurate measurement. This paper has shown that an intelligent processing method based on oscillating second-order systems working on limited time interval can differentiate between large-step inputs (which are active on the whole limited time interval) and short-step inputs (which are active on a time interval shorter than the limited working period). Some resonance aspects (appearing when the input frequency is close to the working frequency of the oscillating second-order system) were also presented. Possible applications could be represented by processing the electric signal generated by transducers [9] and by advanced modeling of traffic network [10].

Acknowledgments

This study was performed by first author (from FILS Department, Politehnica University, Bucharest) during an internship at Corner Soft Technologies, Bucharest (research coordinated by the second author).

References

[1] C. Toma, "An extension of the notion of observability at filtering and sampling devices," in *Proceedings of the International Symposium on Signals, Circuits and Systems (ISSCS '05)*, p. 233, Iasi, Romania, July 2005.

[2] Y. Sun and J. K. Fidler, "Synthesis and performance analysis of universal minimum component integrator-based IFLF OTA-grounded capacitor filter," *IEE Proceedings: Circuits, Devices and Systems*, vol. 143, no. 2, pp. 107–114, 1996.

[3] S. L. Smith and E. Sánchez Sinencio, "Low voltage integrators for high frequency cmos filters using current mode techniques," *IEEE Transactions on Circuits and Systems II*, vol. 43, no. 1, pp. 39–48, 1996.

[4] C. Cattani and J. Rushchitsky, *Wavelet and Wave Analysis as applied to Materials with Micro or Nanostructure*, vol. 74 of *Series on Advances in Mathematics for Applied Sciences*, World Scientific, Singapore, 2007.

[5] C. Cattani, "Connection coefficients of Shannon wavelets," *Mathematical Modelling and Analysis*, vol. 11, no. 2, pp. 117–132, 2006.

[6] G. Toma, "Practical test-functions generated by computer algorithms," in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA '05)*, vol. 3482 of *Lecture Notes in Computer Science*, pp. 576–584, Singapore, May 2005.

[7] S. Pusca, "Invariance properties of practical test-functions used for generating asymmetrical pulses," in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA '06)*, vol. 3980 of *Lecture Notes in Computer Science*, pp. 763–770, Glasgow, UK, May 2006.

[8] A. Toma, S. Pusca, and C. Morarescu, "Spatial aspects of interaction between high-energy pulses and waves considered as suddenly emerging phenomena," in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA '06)*, vol. 3980 of *Lecture Notes in Computer Science*, pp. 839–846, Glasgow, UK, May 2006.

[9] E. Smeu, "Fast photodetectors speed-related problems," *Romanian Journal of Optoelectronics*, vol. 11, no. 1, p. 63, 2003.

[10] M. Li and S. C. Lim, "Modeling network traffic using generalized Cauchy process," *Physica A*, vol. 387, no. 11, pp. 2584–2594, 2008.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 689074, 11 pages
doi:10.1155/2008/689074

Research Article

On Nonperturbative Techniques for Thermal Radiation Effect on Natural Convection past a Vertical Plate Embedded in a Saturated Porous Medium

O. D. Makinde[1] and R. J. Moitsheki[2]

[1] *Faculty of Engineering, Cape Peninsula University of Technology, P.O. Box 1906, Bellville 7535, South Africa*
[2] *School of Computational and Applied Mathematics, University of the Witwatersrand, Private Bag 3, Wits 2050, South Africa*

Correspondence should be addressed to O. D. Makinde, makinded@cput.ac.za

Received 16 March 2008; Revised 30 May 2008; Accepted 24 June 2008

Recommended by Carlo Cattani

In this article, the heat transfer characteristics of natural convection about a vertical permeable flat surface embedded in a saturated porous medium are studied by taking into account the thermal radiation effect. The plate is assumed to have a power-law temperature distribution. Similarity variables are employed in order to transform the governing partial differential equations into a nonlinear ordinary differential equation. Both Adomian decomposition method (ADM) and He's variational iteration method (VIM) coupled with Padé approximation technique are implemented to solve the reduced system. Comparisons with previously published works are performed, and excellent agreement between the results is obtained.

1. Introduction

Heat transfer from different geometrics embedded in porous media has many engineering and geophysical applications such as geothermal reservoirs, drying of porous solids, thermal insulation, enhanced oil recovery, packed-bed catalytic reactors, cooling of nuclear reactors, and underground energy transport [1]. Nakayama and Koyama [2] studied free convection over a vertical flat plate embedded in a thermally stratified porous medium by exploiting the similarity transformation procedure. Cheng and Minkowycz [3] studied the steady free convection about a vertical plate embedded in a porous media using the boundary layer assumptions and Darcy model by the similarity method. Cheng [4] extended the work by studying the effect of lateral mass flux with prescribed temperature and velocity as power law on the vertical surface. Other investigators [5–8] studied some similar porous medium cases using Darcy and Boussinesq approximations with different power-law velocity and temperature variations at the boundaries.

Meanwhile, the boundary layer equations for free-convective flow through a porous medium constitute a nonlinear problem. The theory of nonlinear differential equations is quite elaborate and their solutions are of practical relevance in the engineering sciences. Several numerical approaches have been developed in the last few decades (e.g., finite differences, spectral method, shooting method, etc.) to tackle this problem. More recently, the ideas of classical analytical methods have experienced a revival in connection with the proposition of novel hybrid numerical-analytical schemes for nonlinear differential equations. Among such trends are Adomian decomposition method (ADM) [9–12] and He's variational iteration method (VIM) [13–16] coupled with Padé approximation method [17] especially when dealing with boundary value problems [18]. These techniques, over the last few years, have proved themselves as a powerful tool and a potential alternative to traditional numerical techniques in various applications in science and engineering. This seminumerical approach is also extremely useful in the validation of purely numerical schemes.

The aim of the present work is to construct a nonperturbative solution for natural convection boundary layer flow through a porous medium on an unbound domain in the presence of radiation using both ADM and VIM coupled with Padé approximation technique. The chief merit of the methods is that they are capable of greatly reducing the size of computation work while still maintaining accuracy of the numerical solution. However, VIM gives successive approximations of high accuracy of the solution and VIM does not require specific treatments as in ADM for nonlinear terms. Both numerical and graphical results are presented and discussed quantitatively with respect to various parameters embedded in the problem.

2. Mathematical formulation

We consider the steady two-dimensional flow of an incompressible viscous fluid induced by a heated vertical plate embedded in a homogeneous porous medium of uniform ambient temperature T_∞. The fluid is assumed to be Newtonian, and a constant fluid suction or blowing is imposed at the plate surface. Under Darcy and Boussinesq approximations, the governing boundary layer equations for this problem can be written as [5, 6]

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \tag{2.1}$$

$$\frac{\partial u}{\partial y} = \frac{gK\beta}{\upsilon}\frac{\partial T}{\partial y}, \tag{2.2}$$

$$\rho c_p\left(u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y}\right) = k\frac{\partial^2 T}{\partial y^2} - \frac{\partial q_r}{\partial y}. \tag{2.3}$$

Here, u, v are the velocity components along x, y coordinates, ρ the fluid density, T the temperature of the fluid, c_p specific heat at constant pressure, k the thermal conductivity, υ kinematic viscosity, g the gravitational acceleration, K permeability of the porous medium, and β thermal expansion coefficient. It is also assumed that the temperature distribution of the plate is governed by the power law $T_w(x) = T_\infty + Ax^\lambda$, where A is a constant > 0 for heated plate. Using Roseland approximation [7, 19], we takethe radiative heat flux as

$$q_r = -\frac{4\sigma}{3\gamma}\frac{\partial T^4}{\partial y}, \tag{2.4}$$

where σ is the Stefan-Boltzmann constant and γ the mean absorption coefficient. Assume that the temperature differences within the flow are sufficiently small such that T^4 may expressed as a linear function of temperature

$$T^4 \approx 4T_\infty^3 T - 3T_\infty^4. \tag{2.5}$$

The boundary conditions are given by

$$\begin{aligned} T(x,0) &= T_w(x), \qquad v(x,0) = V(x), \\ T(x,\infty) &= T_\infty, \qquad u(x,\infty) = 0. \end{aligned} \tag{2.6}$$

We introduced the following similarity variables and parameters [5, 8]:

$$\begin{gathered} \Psi = \alpha Ra_x F(\eta), \quad Ra_x = \frac{gK\beta(T_w - T_\infty)x}{\upsilon\alpha}, \; T = T_\infty + Ax^\lambda\theta(\eta), \\ N = \frac{16\sigma T_\infty^3}{3\gamma k}, \quad \eta = \left(\frac{y}{x}\right)Ra_x^{1/2}, \quad \alpha = \frac{k}{\rho c_p}, \quad \theta = \frac{T - T_\infty}{T_w - T_\infty}, \end{gathered} \tag{2.7}$$

where Ra_x is the modified local Rayleigh number. The continuity equation (2.1) is satisfied by the stream function $\Psi(x,y)$ defined by

$$u = \frac{\partial\Psi}{\partial y} = \left(\frac{\alpha}{x}\right)Ra_x F'(\eta), \qquad v = -\frac{\partial\Psi}{\partial x} = -\left(\frac{\alpha}{2x}\right)Ra_x^{1/2}\left[(\lambda+1)F + (\lambda-1)F'\right], \tag{2.8}$$

and (2.2) and (2.3) become

$$F'' = \theta', \qquad \theta'' + \frac{\lambda+1}{2(N+1)}F\theta' - \frac{\lambda}{N+1}F'\theta = 0, \tag{2.9}$$

where the primes denote differentiation with respect to η, N is the Radiation parameter, and λ is the temperature exponent. In view of (2.7), the boundary conditions (2.6) transform into

$$\theta(0) = 1, \qquad \theta(\infty) = 0, \qquad F(0) = m, \qquad F'(\infty) = 0. \tag{2.10}$$

The suction or injection speed at the plate surface becomes

$$v(x,0) = -\left(\frac{\alpha}{2x}\right)Ra_x^{1/2}(\lambda+1)F(0), \tag{2.11}$$

where $m = F(0)$ is the suction or injection parameter according to $m > 0$ or $m < 0$, respectively. The entrainment velocity of the fluid is given by

$$v(x, \infty) = -\left(\frac{\alpha}{2x}\right) Ra_x^{1/2} (\lambda + 1) F(\infty). \tag{2.12}$$

Equation (2.9) together with the boundary conditions (2.10) can be easily reduced to give

$$F''' + \frac{\lambda + 1}{2(N+1)} F'' F - \frac{\lambda}{N+1} F'^2 = 0, \tag{2.13}$$

with

$$F'(0) = 1, \qquad F(0) = m, \qquad F'(\infty) = 0, \tag{2.14}$$

since it is very obvious from (2.9) and (2.10) that $F' = \theta$ (i.e., the vertical velocity and the temperature profiles are identical). The local surface heat flux can be expressed as a function of the local Rayleigh and Nusselt numbers as

$$Nu_x Ra_x^{-1/2} = -\theta'(0). \tag{2.15}$$

3. Adomian decomposition method

In order to explicitly construct approximate nonperturbative solutions of the problem described by (2.13) and (2.14), Adomian decomposition method well addressed in [9–11] is employed and implemented in Maple (a symbolic algebra package). We rewrite (2.13) in the form

$$L_\eta F = \frac{\lambda}{(N+1)} (F_\eta)^2 - \frac{(\lambda + 1)}{2(N+1)} F F_{\eta\eta}, \tag{3.1}$$

where the subscript η represents differentiation with respect to η and the differential operator employs the first three derivatives in the form $L_\eta = d^3/d\eta^3$. The inverse operator L_η^{-1} is considered a threefold integral operator defined by

$$L_\eta^{-1} = \int_0^\eta \int_0^\eta \int_0^\eta (\cdot) d\eta d\eta d\eta. \tag{3.2}$$

Applying L_η^{-1} to both sides of (3.1), using the boundary conditions in (2.14), we obtain

$$F(\eta) = m + \eta + b\frac{\eta^2}{2} + a_1 L_\eta^{-1}(F_\eta^2) - a_2 L_\eta^{-1}(F F_{\eta\eta}), \tag{3.3}$$

where $a_1 = \lambda/(N+1)$, $a_2 = (\lambda+1)/2(N+1)$, and $b = F''(0)$ is to be determined from the boundary condition at infinity in (2.14). As usual in Adomian decomposition method, the solution of (3.3) is approximated as an infinite series

$$F(\eta) = \sum_{j=0}^{\infty} F_j, \tag{3.4}$$

and the nonlinear terms are decomposed as

$$F_\eta^2 = \sum_{j=0}^{\infty} H_j, \qquad FF_{\eta\eta} = \sum_{j=0}^{\infty} G_j, \tag{3.5}$$

where H_j, G_j, are polynomials (called Adomian polynomials) given by

$$\begin{gathered} H_j = \frac{1}{j!}\frac{d^j}{dS^j}\left[\left(\sum_{i=0}^{\infty} F_{\eta i} S^i\right)^2\right]_{S=0}, \\ G_j = \frac{1}{j!}\frac{d^j}{dS^j}\left[\left(\sum_{i=0}^{\infty} F_i S^i\right)\left(\sum_{i=0}^{\infty} F_{\eta\eta i} S^i\right)\right]_{S=0}. \end{gathered} \tag{3.6}$$

Thus, we can identify

$$\begin{gathered} F_0 = m + \eta, \qquad F_1 = \frac{b\eta^2}{2} + a_1 L_\eta^{-1}(H_0) - a_2 L_\eta^{-1}(G_0), \\ F_{j+1} = a_1 L_\eta^{-1}(H_j) - a_2 L_\eta^{-1}(G_j), \quad \text{for } j \geq 1. \end{gathered} \tag{3.7}$$

Using Maple, we obtained a few terms approximation to the solution as

$$\begin{aligned} F_1 &= \frac{1}{2}b\eta^2 + \frac{(1/6)\lambda\eta^3}{N+1}, \\ F_2 &= \left(\frac{1}{60}\frac{\lambda^2}{(N+1)^2} - \frac{1}{60}\frac{(1+\lambda)\lambda}{(2N+2)(N+1)}\right)\eta^5 \\ &\quad + \left(\frac{1}{12}\frac{\lambda b}{N+1} - \frac{(1+\lambda)((1/24)(m\lambda/(N+1)) + (1/24)b)}{2N+2}\right)\eta^4 - \frac{1}{6}\frac{(1+\lambda)mb\eta^3}{2N+2} \end{aligned} \tag{3.8}$$

and so on. Substituting (3.7) into (3.4), we obtain

$$\begin{aligned}F(\eta) = {} & m + \eta + \frac{1}{2}b\eta^2 - \frac{1}{12}\frac{(bm + bm\lambda - 2\lambda)\eta^3}{N+1} \\ & + \frac{1}{96}\frac{1}{(N+1)^2}((-2bN - 2b + 6b\lambda N + 6b\lambda + bm^2 + 2bm^2\lambda - 2m\lambda + bm^2\lambda^2 - 2m\lambda^2)\eta^4) \\ & + \frac{1}{480}\frac{1}{(N+1)^2}((-2b^2N - 2b^2 + 6b^2\lambda N + 6b^2\lambda - 2bm\lambda - 5bm\lambda^2 + 4\lambda^2 + 3bm - 4\lambda)\eta^5) \\ & + \mathrm{O}(\eta^6),\end{aligned} \tag{3.9}$$

where other terms up to $\mathrm{O}(\eta^{13})$ were derived. Let $W_L = \sum_{j=0}^{L} F_j$ represent the decomposition series partial sum obtained, then $F(\eta) = \lim_{L\to\infty}(W_l)$.

4. He's variational iteration method

In 1978, Inokuti et al. [20] proposed a general Lagrange multiplier method to solve nonlinear problems, which was first proposed to solve problems in quantum mechanics. The modified method, or variational iteration method (VIM) proposed by He [13–16], has been shown to solve effectively, easily, and accurately a large class of nonlinear problems with approximations converging rapidly to accurate solutions. To illustrate the basic idea of the method, we consider the general nonlinear system

$$L[F(\eta)] + N[F(\eta)] = g(\eta), \tag{4.1}$$

where L is a linear operator, N is a nonlinear operator, and $g(\eta)$ is a given continuous function. The basic character of the method is to construct a correction functional for the system, which reads

$$F_{n+1}(\eta) = F_n(\eta) + \int_0^\eta B(s)\left[LF_n(s) + N\tilde{F}_n(s) - g(s)\right]ds, \tag{4.2}$$

where B is a Lagrange multiplier which can be identified optimally via variational theory, F_n is the nth approximate solution, and $\tilde{F}_n$ denotes a restricted variation (i.e., $\delta\tilde{F}_n = 0$). This technique provides a sequence of functions which converges to the exact solution of the problem. The initial values $F(0)$, $F'(0)$, and $F''(0)$ are usually used for selecting the zeroth approximation F_0. Consequently, the exact solution may be obtained by using $F(\eta) = \lim_{n\to\infty} F_n$.

In what follows, we will apply the VIM for the problem in (2.13) to illustrate the strength of the method. The correction functional for (2.13) reads

$$F_{n+1}(\eta) = F_n(\eta) + \int_0^\eta B(s)\left[F_n''' + \frac{\lambda+1}{2(N+1)}\tilde{F}_n''\tilde{F}_n - \frac{\lambda}{N+1}\tilde{F}_n'^2\right]ds. \tag{4.3}$$

Making the above correction functional stationary with respect to F_n yields the stationary conditions (Euler equations)

$$\frac{\partial^3 B}{\partial s^3} = 0, \qquad 1 + \left.\frac{\partial^2 B}{\partial s^2}\right|_{s=\eta} = 0, \qquad \left.\frac{\partial B}{\partial s}\right|_{s=\eta} = 0, \qquad B(s)|_{s=\eta} = 0. \tag{4.4}$$

Solving the above equations results in $B = -(1/2)(s-\eta)^2$, and (4.3) then becomes

$$F_{n+1}(\eta) = F_n(\eta) - \frac{1}{2}\int_0^{\eta}(s-\eta)^2\left[F_n'''(s) + \frac{\lambda+1}{2(N+1)}F_n''(s)F_n(s) - \frac{\lambda}{N+1}F_n'^2(s)\right]ds. \tag{4.5}$$

We select the initial value $F_0(\eta) = m + \eta + (1/2)b\eta^2$ by using the conditions in (2.14), where $b = F''(0)$ is to be determined from the boundary condition at infinity in (2.14). Using (4.5), we obtain the next successive approximation as

$$\begin{aligned} F_1(\eta) = {} & \left(-\frac{1}{120}\frac{(1+\lambda)b^2}{2N+2} + \frac{1}{60}\frac{\lambda b^2}{N+1}\right)\eta^5 + \left(-\frac{1}{24}\frac{(1+\lambda)b}{2N+2} + \frac{1}{12}\frac{\lambda b}{N+1}\right)\eta^4 \\ & + \left(-\frac{1}{6}\frac{(1+\lambda)bm}{2N+2} + \frac{1}{6}\frac{\lambda}{N+1}\right)\eta^3 + \frac{1}{2}b\eta^2 + \eta + m \end{aligned} \tag{4.6}$$

and after few iterations, we obtain

$$\begin{aligned} F(\eta) = {} & m + \eta + \frac{1}{2}b\eta^2 - \frac{1}{12}\frac{(bm + bm\lambda - 2\lambda)\eta^3}{N+1} \\ & + \frac{1}{96}\frac{1}{(N+1)^2}((-2bN - 2b + 6b\lambda N + 6b\lambda + bm^2 + 2bm^2\lambda - 2m\lambda + bm^2\lambda^2 - 2m\lambda^2)\eta^4) \\ & + \frac{1}{480}\frac{1}{(N+1)^2}((-2b^2N - 2b^2 + 6b^2\lambda N + 6b^2\lambda - 2bm\lambda - 5bm\lambda^2 + 4\lambda^2 + 3bm - 4\lambda)\eta^5) \\ & + \mathrm{O}(\eta^6) \end{aligned} \tag{4.7}$$

and $F(\eta) = \lim_{n\to\infty} F_n$.

5. Padé approximation technique

It is now well known that Padé approximants [17] have the advantage of manipulating the polynomial approximation into rational functions of polynomials. By this manipulation we gain more information about the mathematical behavior of the solution. In addition, power series is not useful for large values of η. Boyd [18] and others have formally shown that power series in isolation are not useful to handle boundary value problems. This can be attributed to the possibility that the radius of convergence may not be sufficiently large to contain the boundaries of the domain. It is therefore essential to combine the series solution, obtained by the ADM and VIM or any series solution methods, with the Padé approximants to provide

Table 1: Comparison of the values of $F''(0)$ with previous results for $N = 0$, $\lambda = 1$.

m	Postelnicu et al. [6]	Ali [5]	Present results (Padé [3/3])	Present results (Padé [6/6])
−1.0	−0.6180	−0.61803	−0.61728	−0.61803
−0.8	−0.6770	−0.67703	−0.67516	−0.67703
−0.6	−0.7440	—	−0.74412	−0.74404
−0.4	−0.8198	−0.81980	−0.82101	−0.81982
−0.2	−0.9050	—	−0.90513	−0.90499
0.0	−1.0000	−1.00000	−0.99998	−1.00000
0.2	−1.1049	—	−1.10524	−1.10497
0.4	−1.2198	—	−1.22781	−1.21976
0.6	−1.3440	—	−1.34462	−1.34390
0.8	−1.4770	—	−1.47789	−1.47701
1.0	−1.6180	−1.61803	−1.62351	−1.61803

Table 2: Numerical values of $F''(0)$ for $N > 0$ using Padé approximants [6/6].

N	λ	m	$F''(0)$
1.0	1.0	1.0	−0.999822
2.0	1.0	1.0	−0.767530
3.0	1.0	1.0	−0.640354
4.0	1.0	1.0	−0.593069
5.0	1.0	1.0	−0.528624
1.0	1.0	0.0	−0.707100
1.0	1.0	0.5	−0.843048
1.0	1.0	−0.5	−0.593084
1.0	1.0	−1.0	−0.500004

an effective tool to handle boundary value problems on an infinite or semi-infinite domain. Recall that the Padé approximants can be easily evaluated by using built-in function in a symbolic computational package such as Maple. The essential behavior of the solution will be addressed by using several diagonal Padé approximants of different degrees. Furthermore, the undetermined value of $b = F''(0)$ is calculated from the boundary condition at infinity in (2.14). The difficulty at infinity is overcome by employing the diagonal Padé approximants [10, 11, 18] that approximate $F'(\eta)$ using $W'_L(\eta)$. For instance, the series is transformed into diagonal Padé approximants as follows:

$$W'_{L[M,M]}(\eta) = \frac{\sum_{i=0}^{M} h_i \eta^i}{\sum_{i=0}^{M} g_i \eta^i}, \tag{5.1}$$

where $P = 2(M + 1)$ is the order of the series required for each approximant. In the Maple environment, the simultaneous evaluation of $\lim_{\eta\to\infty} W'_{L[M/M]}(\eta) = 0$ for $M = 2, 3, 4, \ldots$ in (3.9) gives the numerical results for $b = F''(0)$ as shown in Tables 1 and 2.

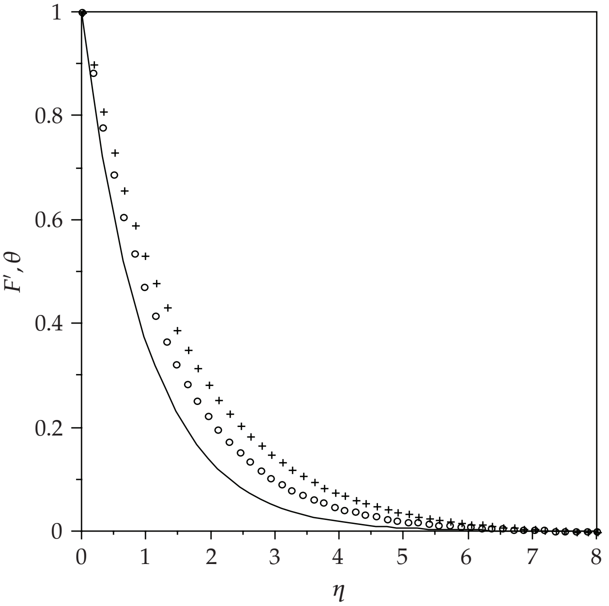


Figure 1: Vertical velocity or temperature profiles for $\lambda = 1$; $m = 1$; $N = 1$ (solid line), $N = 2$ (circles), $N = 3$ (plus signs).

6. Numerical results and discussion

The governing equation (2.13) subject to the boundary conditions (2.14) is solved using both ADM and VIM together with Padé technique as described in Sections 3–5. Solutions are obtained for the plate temperature with uniform lateral mass flux ($\lambda = 1$) controlled by the suction/injection parameter m and radiation parameter N as shown in Tables 1 and 2 and Figures 1 and 2.

The results presented in Table 1 are in good agreement with those given by Postelnicu et al. [6] and Ali [5] who solved numerically the case of permeable surface without considering the thermal radiation effect. In Table 2, we observed that the local surface heat flux rate decreases with increasing values of radiation parameter. Figures 1 and 2 confirm the exponential decay velocity $F'(\eta)$ or temperature $\theta(\eta)$ profiles across the boundary layers [5–8]. As mentioned earlier, suction corresponds to $m > 0$, injection to $m < 0$, and $m = 0$ to impermeable plate. Therefore, it is clear from Figure 1 that suction reduces the boundary layer thickness sharply as seen for $m = 1$ while injection increases it as for $m = -1$; however, the surface heat flow is always positive regardless of the sign of m where the heat is directed from the plate to the porous medium. Figure 2 shows that the fluid velocity and temperature increase as the radiation parameter N increases. This can be explained by the fact that the effect of radiation N is to increase the rate of energy transport to the fluid and accordingly to increase the fluid temperature.

7. Conclusions

We employed both ADM and VIM to compute a nonperturbative solution for thermal radiation effect on natural convection boundary layer flow past a vertical plate embedded in a saturated porous medium. The results demonstrate the reliability and the efficiency

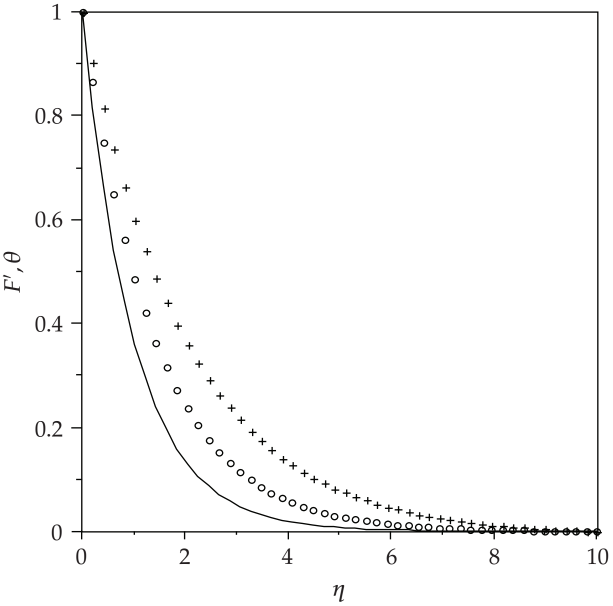


Figure 2: Vertical velocity or temperature profiles for $N = 1$; $\lambda = 1$; $m = 1$ (solid line), $m = 0$ (circles), $m = -1$ (plus signs).

of both methods in an unbounded domain. The two methods are powerful and efficient in obtaining approximations of higher accuracy and closed-form solutions if existing. However, He's variational iteration method gives several successive approximations through using the iteration of the correction functional and Adomian decomposition method provides the components of the exact solution that will be added to get the series solution. Moreover, the VIM requires the evaluation of the Lagrangian multiplier whereas ADM requires the evaluation of the Adomian polynomials that mostly require tedious algebraic calculations.

Acknowledgment

The authors would like to thank the National Research Foundation of South Africa Thuthuka programme for financial support.

References

[1] D. A. Nield and A. Bejan, *Convection in Porous Media*, chapter 5, Springer, New York, NY, USA, 2nd edition, 1999.

[2] A. Nakayama and H. Koyama, "Free convective heat transfer over a non-isothermal body of arbitrary shape embedded in a fluid saturated porous medium," *Journal of Heat Transfer*, vol. 109, no. 1, pp. 125–130, 1987.

[3] P. Cheng and W. J. Minkowycz, "Free convection about a vertical flat plate embedded in a porous medium with application to heat transfer from a dike," *Journal of Geophysical Research*, vol. 82, no. 14, pp. 2040–2044, 1977.

[4] P. Cheng, "The influence of lateral mass flux on free convection boundary layers in a saturated porous medium," *International Journal of Heat and Mass Transfer*, vol. 20, no. 3, pp. 201–206, 1977.

[5] M. E. Ali, "The effect of lateral mass flux on the natural convection boundary layers induced by a heated vertical plate embedded in a saturated porous medium with internal heat generation," *International Journal of Thermal Sciences*, vol. 46, no. 2, pp. 157–163, 2007.

[6] A. Postelnicu, T. Groşan, and I. Pop, "Free convection boundary-layer over a vertical permeable flat plate in a porous medium with internal heat generation," *International Communications in Heat and Mass Transfer*, vol. 27, no. 5, pp. 729–738, 2000.

[7] O. D. Makinde, "Free convection flow with thermal radiation and mass transfer past a moving vertical porous plate," *International Communications in Heat and Mass Transfer*, vol. 32, no. 10, pp. 1411–1419, 2005.

[8] T. Groşan and I. Pop, "Free convection of non-Newtonian fluids over a vertical surface in a porous medium with internal heat generation," *International Journal of Applied Mechanics and Engineering*, vol. 7, no. 2, pp. 401–407, 2002.

[9] G. Adomian, *Solving Frontier Problems of Physics: The Decomposition Method*, vol. 60 of *Fundamental Theories of Physics*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[10] A.-M. Wazwaz, "The modified decomposition method and Padé approximants for a boundary layer equation in unbounded domain," *Applied Mathematics and Computation*, vol. 177, no. 2, pp. 737–744, 2006.

[11] S. A. Kechil and I. Hashim, "Non-perturbative solution of free-convective boundary-layer equation by Adomian decomposition method," *Physics Letters A*, vol. 363, no. 1-2, pp. 110–114, 2007.

[12] T. A. Abassy, M. A. El-Tawil, and H. K. Saleh, "The solution of Burgers' and good Boussinesq equations using ADM-Padé technique," *Chaos, Solitons &amp; Fractals*, vol. 32, no. 3, pp. 1008–1026, 2007.

[13] J.-H. He and X.-H. Wu, "Variational iteration method: new development and applications," *Computers &amp; Mathematics with Applications*, vol. 54, no. 7-8, pp. 881–894, 2007.

[14] J.-H. He, "Variational iteration method—some recent results and new interpretations," *Journal of Computational and Applied Mathematics*, vol. 207, no. 1, pp. 3–17, 2007.

[15] J.-H. He and X.-H. Wu, "Construction of solitary solution and compacton-like solution by variational iteration method," *Chaos, Solitons &amp; Fractals*, vol. 29, no. 1, pp. 108–113, 2006.

[16] J.-H. He, "Some asymptotic methods for strongly nonlinear equations," *International Journal of Modern Physics B*, vol. 20, no. 10, pp. 1141–1199, 2006.

[17] G. A. Baker Jr., *Essentials of Padé Approximants*, Academic Press, London, UK, 1975.

[18] J. Boyd, "Padé approximant algorithm for solving nonlinear ordinary differential equation boundary value problems on an unbounded domain," *Computers in Physics*, vol. 11, no. 3, pp. 299–303, 1997.

[19] M. A. Seddeek, "Thermal radiation and buoyancy effects on MHD free convection heat generation flow over an accelerating permeable surface with temperature dependent viscosity," *Canadian Journal of Physics*, vol. 79, no. 4, pp. 725–732, 2001.

[20] M. Inokuti, H. Sekine, and T. Mura, "General use of the Lagrange multiplier in nonlinear mathematical physics," in *Variational Method in the Mechanics of Solids*, S. Nemat-Nassed, Ed., pp. 156–162, Pergamon press, New York, NY, USA, 1978.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 410156, 20 pages
doi:10.1155/2008/410156

Research Article

Relativistic Short Range Phenomena and Space-Time Aspects of Pulse Measurements

Ezzat G. Bakhoum[1] and Cristian Toma[2]

[1] *Department of Electrical and Computer Engineering, University of West Florida,
11000 University Parkway, Pensacola, FL 32514, USA*
[2] *Faculty of Applied Sciences, Politechnica University, Hagi-Ghita 81, 060032 Bucharest, Romania*

Correspondence should be addressed to Ezzat G. Bakhoum, ebakhoum@uwf.edu

Received 22 April 2008; Accepted 5 May 2008

Recommended by Carlo Cattani

Particle physics is increasingly being linked to engineering applications via electron microscopy, nuclear instrumentation, and numerous other applications. It is well known that relativistic particle equations notoriously fail over very short space-time intervals. This paper introduces new versions of Dirac's equation and of the Klein-Gordon equation that are suitable for short-range phenomena. Another objective of the paper is to demonstrate that pulse measurement methods that are based on the wave nature of matter do not necessarily correlate with physical definitions that are based on the corpuscular nature of particles.

1. Introduction

The theory of special relativity plays a great role in particle physics. Now, particle physics is increasingly being linked to engineering applications, via electron microscopy, superconductivity, nuclear instrumentation, to name a few applications. Since relativistic formulae are at the heart of all such applications, then it becomes important to find ways to perform numerical computations related to localized (short-range) relativistic phenomena. For instance, it is well known that the relativistic version of Schrödinger's equation, namely, the Dirac equation, cannot normally be solved over a short interval because it always predicts that the velocity of the electron is equal to c, or the speed of light. In applications such as electron microscopy, it becomes therefore usually necessary to abandon the relativistic formulae and rely solely on the classical theory of electromagnetism. It is therefore clear that there is a need at the present time to formulate the Dirac and other relativistic equations in a manner that allows the computation of short-range phenomena. This is the first objective of this paper.

The second objective of this paper is to show that space-time measurements on closed-loop trajectories in special relativity and noncommutative properties of operators in quantum physics require a more rigorous definition of the method of measurement of interaction phenomena. The use of the least action principle, for instance, implies some logic definitions for measuring methods that are based on waves and for measuring methods that are based on the corpuscular aspects of matter. When measurement is applied to pulses, those logic definitions include considerations about a possible memory of previous measurements (space-time operators). Accordingly, a distinction exists between the set of existing space-time intervals and the set of measured space-time intervals (established using wave measurement methods and defined within limited space-time intervals).

2. Relativistic short-range electron equation

In this section, we will develop a version of Dirac's equation that is suitable for pulsed, short-range electron beams. We will rely on the recently introduced mass-energy equivalence relation $H = mv^2$ [1] (where H is the total energy of the electron and v is its velocity), which has proved to be effective in explaining short-range phenomena. First, a new Hamiltonian will be obtained. It will be then verified that the new Hamiltonian directly leads to the result that the velocity of the electron must be equal to $\pm v$, which is of course a result that is in sharp contrast with Dirac's result and which does agree with experimental observation. We will also verify that the spin magnetic moment term obtained by Dirac remains unchanged in the present formulation.

2.1. The wave equation

We will begin by describing briefly Dirac's approach for obtaining the relativistic wave equation and then proceed to derive the modified equation and hence the modified Hamiltonian. Dirac considered the mass of the particle as represented by its relativistic expression $m = m_0/\sqrt{1 - v^2/c^2}$. If we square that expression and rearrange the terms, we get

$$m^2c^2 = m^2v^2 + m_0^2c^2. \tag{2.1}$$

Multiplying by c^2, we get

$$m^2c^4 = m^2v^2c^2 + m_0^2c^4. \tag{2.2}$$

But since mc^2 is the total energy according to Einstein, then we have

$$H^2 = p^2c^2 + m_0^2c^4. \tag{2.3}$$

Hence,

$$H = c\sqrt{p^2 + m_0^2c^2}. \tag{2.4}$$

Since the term p^2 can be written as $\sum_r p_r^2$, where p_r is a one-dimensional momentum component and $r = 1, 2, 3$, we finally have

$$H = c\sqrt{\sum_r p_r^2 + m_0^2c^2}. \tag{2.5}$$

This was Dirac's total energy equation and was subsequently used to obtain the relativistic wave equation. To obtain the modified wave equation, we now proceed to multiply (2.1) by v^2, getting

$$m^2c^2v^2 = m^2v^4 + m_0^2c^2v^2. \tag{2.6}$$

Using $H = mv^2$ as the total energy of the particle, we have, from the above expression,

$$H^2 = p^2c^2 - m_0^2c^2v^2. \tag{2.7}$$

Now, since $v^2 = \sum_r v_r^2$, where v_r is a one-dimensional velocity component, (2.7) can equivalently be written as

$$H = c\sqrt{\sum_r p_r^2 - m_0^2 \sum_r v_r^2}. \tag{2.8}$$

Equation (2.8) can be further simplified by noting that $v_r = p_r/m$. We finally have

$$H = c\sqrt{\left(1 - \frac{m_0^2}{m^2}\right)\sum_r p_r^2}. \tag{2.9}$$

Following Dirac's approach, if we let $\vec{p}_0$ be a vector defined as $\vec{p}_0 = \vec{H}/c$, where $\vec{H}$ may be Hamiltonian of the form $\vec{H} = (H, 0, 0)$, we will seek a wave equation that is linear in $\vec{p}_0$. We will take an equation of the most simple, basic form

$$\left(\vec{p}_0 - \sum_r \vec{p}_r [\alpha_r]\right)\psi = 0. \tag{2.10}$$

This form can be sufficient without any additional terms if we do not impose any restrictions on the matrices $[\alpha_r]$. Dirac found that such matrices must be noncommuting, but it is obvious here that such matrices must also contain mass terms. Multiplying (2.10) by the vector $(\vec{p}_0 + \sum_r \vec{p}_r [\alpha_r])$, we get

$$p_0^2 - \left(\sum_r \vec{p}_r [\alpha_r]\right)^2 = 0. \tag{2.11}$$

Comparing this last expression with (2.9), we conclude that

$$\left(1 - \frac{m_0^2}{m^2}\right)\sum_r p_r^2 = \left(\sum_r \vec{p}_r [\alpha_r]\right)^2 = \sum_r \vec{p}_r [\alpha_r]^2 \vec{p}_r^{\,T} + \sum_j \sum_k \vec{p}_j [\alpha_j][\alpha_k] \vec{p}_k^{\,T}, \tag{2.12}$$

where $j, k = 1, 2, 3$, and $j \neq k$. Accordingly, the matrices $[\alpha_r]$ must satisfy

$$[\alpha_r] = \pm\sqrt{1 - \frac{m_0^2}{m^2}}\,[\beta_r], \tag{2.13}$$

where $[\beta_r]$ are matrices that must satisfy the following two conditions:

$$[\beta_r]^2 = I, \qquad [\beta_j][\beta_k] + [\beta_k][\beta_j] = 0. \tag{2.14}$$

Examples of such matrices were suggested by Dirac [2]. They might take the following forms among others:

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}. \tag{2.15}$$

(Note that Dirac used 4 × 4 matrices to account for time as the fourth dimension. It was independently confirmed later that the minimum number of dimensions that will satisfy Dirac's theory is in fact four.)

Using the relativistic expression for m, the matrices $[\alpha_r]$ can now be written as

$$[\alpha_r] = \pm\sqrt{1 - \left(1 - \frac{v^2}{c^2}\right)}\,[\beta_r] = \pm\frac{v}{c}\,[\beta_r]. \tag{2.16}$$

Therefore, from (2.12) and (2.16), the vector Hamiltonian can be written as

$$\vec{H} = c\,\vec{p}_0 = c\sum_r \vec{p}_r\,[\alpha_r] \pm v\sum_r \vec{p}_r\,[\beta_r]. \tag{2.17}$$

To check the modified theory, it can be now easily verified that the velocity component $\dot{x}_1$ will be given by

$$\dot{x}_1 = [x_1, \vec{H}] = \pm v. \tag{2.18}$$

Unlike Dirac's result, this result is of course in agreement with experimental observation. It is important to note here that, mathematically, $\dot{x}_1$ is the "expected" value of the velocity. From (2.17), we can also see that the negative energy states are still preserved here.

2.2. Motion of a charged particle in a magnetic field

We now consider the motion of a charged particle in a magnetic field to obtain a formulation for the spin magnetic moment term that must appear in the final Hamiltonian (we assume the absence of an electrostatic field here). In the presence of a magnetic field, the change in the particle momentum Δp that occurs as a result of the interaction with the field is given by [3]

$$\Delta p = \frac{e}{c}\,A, \tag{2.19}$$

where e is the particle charge and A is the magnitude of the vector magnetic potential. Adding that term to the momentum in (2.17) gives the Hamiltonian

$$\vec{H} = \pm v\sum_r \left(\vec{p}_r + \frac{e}{c}\,\vec{A}_r\right)[\beta_r]. \tag{2.20}$$

By squaring (2.20), we get

$$\frac{H^2}{v^2} = \sum_r \left[\left(\vec{p}_r + \frac{e}{c} \vec{A}_r \right) [\beta_r] \right]^2 + \sum_j \sum_k \left(\vec{p}_j + \frac{e}{c} \vec{A}_j \right) [\beta_j] [\beta_k] \left(\vec{p}_k + \frac{e}{c} \vec{A}_k \right)^T . \tag{2.21}$$

It is fairly easy to verify that the second term on the r.h.s. of the above expression must vanish since the $\vec{p}_r$ vectors commute and since the $[\beta_r]$ matrices satisfy condition (2.14). In Dirac's treatment of the subject, he was able to show that the following equation holds:

$$\left[\left(\vec{p}_r + \frac{e}{c} \vec{A}_r \right) [\beta_r] \right]^2 \left(\vec{p}_r + \frac{e}{c} \vec{A}_r \right)^2 + \frac{\hbar e}{c} \| \vec{M} \, [\beta_r] \|, \tag{2.22}$$

where $\vec{M} = \text{curl} \, \vec{A}$ is the magnetic field intensity vector. Equation (2.21) therefore becomes

$$H^2 = v^2 \sum_r \left(\vec{p}_r + \frac{e}{c} \vec{A}_r \right)^2 + v^2 \frac{\hbar e}{c} \sum_r \| \vec{M} \, [\beta_r] \| . \tag{2.23}$$

(Note that (2.23) is a scalar equation.) If we now let $H = mv^2$ and divide both sides of the equation by $2mv^2$, we get

$$\frac{1}{2} mv^2 = \frac{1}{2m} \sum_r \left(\vec{p}_r + \frac{e}{c} \vec{A}_r \right)^2 + \frac{\hbar e}{2mc} \sum_r \| \vec{M} \, [\beta_r] \| . \tag{2.24}$$

If the particle is an electron, then e is a negative quantity and the above equation becomes

$$\frac{1}{2m} \sum_r \left(\vec{p}_r + \frac{e}{c} \vec{A}_r \right)^2 = \frac{1}{2} mv^2 + \frac{\hbar |e|}{2mc} \sum_r \| \vec{M} \, [\beta_r] \| . \tag{2.25}$$

Without the presence of the magnetic field, the l.h.s. of (2.25) is reduced to

$$\frac{1}{2m} \sum_r p_r^2 = \frac{p^2}{2m} . \tag{2.26}$$

This is the same as $1/2 \, mv^2$. We can therefore conclude that the second term on the r.h.s. of (2.25) is the term that represents the interaction of the field with the electron magnetic moment. Hence the quantity $\hbar |e| / 2mc$ is the spin magnetic moment coefficient. In general, we can withdraw here the following two important conclusions: (1) the modified theory fully yielded the classical expression of kinetic energy with the addition of the spin interaction term; and (2) the spin interaction term obtained here is the same as the one obtained by Dirac [2] (which is one Bohr magneton). The second conclusion is a confirmation that this part of Dirac's theory was correct. The first conclusion, however, shows a fact that was not apparent from Dirac's theory. Specifically, when $\vec{M} = 0$ (i.e., when the particle is away from the magnetic field lines), (2.25) becomes

$$\frac{1}{2m} \sum_r \left(\vec{p}_r + \frac{e}{c} \vec{A}_r \right)^2 = \frac{1}{2} mv^2 . \tag{2.27}$$

This is a direct confirmation of the Aharonov-Bohm effect [4, 5]. Clearly, (2.27) shows that the components p_r of the momentum will be altered while the kinetic energy remains constant.

3. Phase and group velocities of short-range electrons

The concepts of the phase velocity and the group velocity are very important concepts that come into play when short-range phenomena are considered. For instance, de Broglie's work predicts that the phase velocity of a matter wave is given by the expression c^2/v, which is a very unrealistic assumption for short-range, slow electrons. We will attempt in this section to give a better explanation for that problem.

First of all, we must realize that there exists a number of phase velocities, not a single phase velocity. Now, it is well known mathematically that each phase velocity $v_p = \omega_i/k_i$ and that the group velocity $v_g = d\omega/dk$ (where ω is the angular frequency and $k = 2\pi/\lambda$ is the propagation constant). As was pointed out in [1], the two fundamental relationships of wave mechanics, $\lambda = h/p$ and $H = h\nu$, together make a *statement* about the total energy of a particle; that is, $H = (p\lambda)\nu = pu$, where u is some velocity. The question here is what is u exactly? Is it a phase velocity or a group velocity? Apart from the fact that $H = pu$ is a *total* energy equation, we must also note, since $H = \hbar\omega$ and $p = \hbar k$, that the equation leads to the relationship $\omega = ku$. Hence we must conclude that

$$\frac{d\omega}{dk} = \frac{\omega}{k} = u. \tag{3.1}$$

This means that the group and the phase velocities are *the same*. This is the conclusion that we must hold as true for short-range phenomena. Let us now attempt to understand the origin of the problem. De Broglie's original derivation of the important relationship $\lambda = h/p$ can be found in a number of standard references (see, e.g., [6]). Amazingly, as we will conclude, while the formula was correct, the approach that was used to derive it *was not*.

De Broglie started by assuming a wave function that describes a stationary particle of the form $\psi' = \exp(i\omega' t')$. By using the Lorentz transformation of time $t' = \gamma(t - vx/c^2)$, then $\psi' = \exp(i\gamma\omega'[t - vx/c^2])$. Since this equation (in principle) is a traveling wave equation, de Broglie then concluded that the quantity c^2/v must represent the velocity of the wave in the observer frame. The rest of the derivation that leads to the formula $\lambda = h/p$ is then straightforward and consists of letting $H = h\nu = mc^2$ and substituting the product $\lambda\nu$ for the quantity c^2/v. As it is well known historically [7, 8], de Broglie later offered the hypothesis that c^2/v is only a "phase" velocity and that the real, or "group" velocity is actually v so that the particle and its associated wave would not part company. However, as we indicated, the problem with such a hypothesis is that it directly contradicts the simple conclusion in (3.1).

Let us try to understand the problem with the above approach that led to the indicated contradictions. The Lorentz transformation of time $t' = \gamma(t - vx/c^2)$, which includes the coordinate x, strictly assumes that "x" is only one geometrical point. From the viewpoint of a stationary observer, a traveling wave, in the observer's frame, cannot be described by one "x" coordinate. The correct approach for including a traveling wave within the relativistic transformations is to assume first that the "x" coordinate is equal to zero (and hence the time transformation will be $t' = \gamma t$) and then write a *true* traveling wave equation in the observer frame, that is,

$$\psi = \exp i(kx - \omega t). \tag{3.2}$$

This was indeed the approach that was taken by Shrödinger and certainly this explains why Shrödinger's equation has been unquestionably successful. Now, by noting that $k = 2\pi/\lambda$ and $\omega = 2\pi\nu$, ψ can be written as

$$\psi = \exp i\left(\frac{2\pi}{\lambda}x - \omega t\right) = \exp i\omega\left(\frac{2\pi}{\lambda}\frac{x}{2\pi\nu} - t\right) = \exp i\omega\left(\frac{x}{\lambda\nu} - t\right). \tag{3.3}$$

Assume first that the particle is moving with a velocity $v \ll c$ so that the relativistic effects can be ignored. In this case, ordinary (nonrelativistic) wave mechanics state that $\lambda\nu = v$, or the wave velocity. Now, if the relativistic effect is to be included, then the wavelength λ becomes λ/γ (length contraction) and the frequency ν becomes $\gamma\nu$ (frequency shift). The result therefore is that $\lambda\nu$ is still equal to v. We can see, then, that the flaw in the original approach that led to the result $\lambda\nu = c^2/v$ was the incorrect use of the Lorentz transformation.

If we now follow the rest of de Broglie's derivation, but use $H = mv^2$ instead of mc^2, we have $H = mv^2 = h\nu$, hence

$$p = mv = \frac{h\nu}{v} = \frac{h\nu}{\lambda\nu} = \frac{h}{\lambda}, \tag{3.4}$$

which is of course de Broglie's well-known formula. De Broglie was aware that this relationship can be derived in a number of different ways, and for that reason he raised it to the level of a *postulate*. Concerning the approach that was used in deriving it, however, this is certainly one of the rare cases in science in which an incorrect derivation procedure still led to the correct result.

4. A Klein-Gordon equation and a De Broglie dispersion relation for short-range electrons

In this section, we present derivations for a modified Klein-Gordon equation and a modified de Broglie dispersion relation. The conclusions are: (1) in the case of a massless particle, the dispersion relation is the same as the original one; and (2) in the case of a massive particle, we still conclude that the phase and the group velocities are the same, that is, $v_g = v_p = v$.

4.1. The Klein-Gordon equation

The derivation of the Klein-Gordon equation starts with the usual relativistic expression (see [9])

$$H^2 = p^2c^2 + m_0^2c^4. \tag{4.1}$$

If we now replace H by mc^2 and p by mv, we have

$$m^2c^4 = m^2v^2c^2 + m_0^2c^4. \tag{4.2}$$

If we multiply this expression by v^2/c^2, we get

$$m^2v^2c^2 = m^2v^4 + m_0^2v^2c^2. \tag{4.3}$$

If we now let $H = mv^2$, we finally have

$$H^2 = p^2c^2 - m_0^2v^2c^2. \tag{4.4}$$

This is a modified energy-momentum relationship and was in fact derived previously in [1]. Notice that the quantity $m_0^2v^2 = p^2 - H^2/c^2$. It is therefore a correct representation of the momentum vector p^μ.

To obtain the modified Klein-Gordon equation, we start with the well-known relationship

$$\nabla^2\psi = -\mathbf{k}^2\psi = -\frac{\mathbf{p}^2}{\hbar^2}\psi. \tag{4.5}$$

By substituting from (4.4) into (4.5) we have

$$-\hbar^2\nabla^2\psi = \left(\frac{H^2}{c^2} + m_0^2v^2\right)\psi. \tag{4.6}$$

From Shrödinger's equation we have

$$\frac{\partial^2\psi}{\partial t^2} = -\frac{H^2}{\hbar^2}\psi. \tag{4.7}$$

By substituting from (4.6) into (4.7) we finally get

$$\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} - \nabla^2\psi = \left(\frac{m_0v}{\hbar}\right)^2\psi. \tag{4.8}$$

This is the modified Klein-Gordon equation.

4.2. De Broglie's dispersion relation

In view of (4.7) and (4.5), the modified Klein-Gordon equation can be written as

$$-\frac{1}{c^2}\left(\frac{\omega^2\hbar^2}{\hbar^2}\right)\psi = -\mathbf{k}^2\psi + \left(\frac{m_0v}{\hbar}\right)^2\psi, \quad \text{or,} \quad \omega^2\hbar^2\psi = c^2\hbar^2\mathbf{k}^2\psi - m_0^2c^2v^2\psi. \tag{4.9}$$

Hence, the modified de Broglie wave dispersion relation is

$$\hbar^2\omega^2 = c^2\hbar^2\mathbf{k}^2 - m_0^2c^2v^2. \tag{4.10}$$

For $m_0 = 0$, we can see that the relation becomes $\hbar^2\omega^2 = c^2\hbar^2\mathbf{k}^2$, which is of course the same as in the usual theory.

To obtain the group velocity, $\mathbf{v}_g = d\omega/d\mathbf{k}$, we differentiate the dispersion relation with respect to $\mathbf{k}$, getting (note that only the magnitudes of the vectors $\mathbf{p}$ and $\mathbf{k}$ will be represented)

$$\hbar^2\omega\frac{d\omega}{dk} = c^2\hbar^2k - m_0^2c^2v\frac{dv}{dk}. \tag{4.11}$$

Since $p = mv = \hbar k$ and hence $m(dv/dk) = \hbar$, the above equation becomes

$$\hbar^2\omega\frac{d\omega}{dk} = c^2\hbar^2k - m_0^2c^2\frac{\hbar^2}{m^2}k \tag{4.12}$$

or

$$\omega\frac{d\omega}{dk} = c^2k - \frac{m_0^2}{m^2}c^2k = c^2k\left(1 - \left(1 - \frac{v^2}{c^2}\right)\right) = kv^2. \tag{4.13}$$

Hence,

$$\frac{d\omega}{dk} = \left(\frac{k}{\omega}\right)v^2. \tag{4.14}$$

But since $d\omega/dk = v_g = v$, we then conclude that $\omega/k = v_p = v$. The group and the phase velocities are therefore the same.

5. Logical aspects connected with space-time measurements

After presenting basic aspects in physics from the relativistic point of view, we will present some logical aspects for basic principles in physics (the principle of constant light in vacuum in any reference system and the uncertainty principle in quantum theory). We will show that these principles make use in an implicit manner of terms which are defining also the conclusion. For example, the idea of constant light speed implies the use of a measuring method based on a clocks' synchronization performed using a supposed antecedent light signal transmitted and reflected toward the observer. In a similar manner, the uncertainty principle implies the existence of a measuring method for position or time correlated with a subsequent measurement for momentum or energy (measurements which also make use of position and time). Yet a logic definition of a physics principle cannot be based on the use of the same terms in both sides of it; like in the case of an algebraic calculus, the quantity to be determined must be finally placed in the opposite side of an equality, as related to the already known quantities joined in a mathematical operation. More precisely, we cannot define in a rigorous manner a certain term using the same term in the corresponding definition.

5.1. Logical aspects of light speed constance principle

The constant light speed principle (in vacuum) can be considered under the following form. It exists a quantity *light speed in vacuum* noted as **c**, which is *constant* for any observer inside an inertial reference system.

We can notice at first step that in an implicit manner the previous definition requires the existence of a measuring method for light speed in vacuum; any method for measuring a speed requires the use of time measurements (while $v = \Delta\mathbf{r}/\Delta t$). For our case (special relativity theory), the correspondence of time moments in different reference systems is based on a previous synchronization procedure implying an emission of light from an observer to another and a reflection of this light signal from the other observer to the first one. The reflection moment (considered as synchronization moment ≡ zero moment) is considered by the first observer to take place at the middle of the time interval between the initial emission of light and the return of it. The whole chain implies that the use of a wave light LW appears in the definition of the light speed constance principle (in vacuum) under an explicit form (the notion of light speed), and it appears also under an implicit form (a previous synchronization based on light signals is required). From the formal logic point of view, this represents a contradiction [10]. A first attempt to solve it would be in taking into account the fact that the light speed measurement and the systems synchronization correspond to different time moments (the light wave considered for systems synchronization corresponds to the zero moment of time, while the light wave whose speed is considered in light speed constance principle corresponds to a subsequent moment of time).

However, the use of such a set of different light waves (a light wave whose speed has to be measured and a previous pair of emitted-reflected light wave necessary for the synchronization procedure) implies the use of an extended time interval required by a light speed measurement as

$$T_m = \left[t_0, t_m\right], \tag{5.1}$$

where T_m is the time interval required by a light speed measurement at t_m time moment. But at next step we can notice that a speed corresponds to an almost instant moment of time, being defined as

$$v = \lim_{t \to t_m} \frac{\Delta \mathbf{r}}{\Delta t} = \frac{d\mathbf{r}}{dt}. \tag{5.2}$$

This requires that the time interval required by a speed measurement must be infinitely small. Thus the time interval necessary for light speed measurement can be written as

$$T_m = \left[t_m - \Delta t, t_m\right] \tag{5.3}$$

which implies that the corresponding length interval L_{Tm} is infinitely small

$$L_{Tm} \longrightarrow 0. \tag{5.4}$$

But this is in contradiction with the previous consideration $T_m = [t_0, t_m]$. The corresponding timelength

$$L_{Tm} = t_m - t_0 \gg 0 \tag{5.5}$$

can be much greater than zero. So the contradiction can be easily proved as

$$L_{TM} \longrightarrow 0 \, \text{and in the same time} \, L_{Tm} \gg 0. \tag{5.6}$$

From the intuitive point of view, this means that a light wave emitted in a certain reference system interacts in the most general case only on a limited time interval with another measuring reference system, the use of a previous procedure of emission-reflection for synchronization being impossible in practice. So the solution of such a contradiction (determined by implicit aspects of the terms used in definitions) must be found by taking into consideration other properties of physics entities involved in definition; see also [11].

5.2. Logical aspects of uncertainty principle in quantum mechanics

If we study the uncertainty principle in quantum mechanics, we can notice quite similar aspects. According to this principle, a measurement performed with a greater accuracy upon space or time coordinates for a quantum particle must generate a greater error upon a subsequent measurement for momentum or energy according to

$$\Delta x \Delta P_x \geq \frac{h}{4\pi} \tag{5.7}$$

or

$$\Delta t \Delta E \geq \frac{h}{4\pi}. \tag{5.8}$$

But the existence of a measuring method for position or time is correlated with a subsequent measurement for momentum or energy (measurements which also make use of position and time).

It can be noticed that a term (a space-time measurements) is explained using (in an implicit manner) the same term at a subsequent moment of time. Without being a contradiction (like in the case of light speed constance principle), it still remains a recurrent definition. In the same manner presented for special relativity, we can take into consideration the different moments of time for space-time measurements. Yet the fact that (in an implicit manner) the principle requires the use of a measurement performed at a later time moment generates another logical problem. Can a space-time measurement performed at a certain moment of time be influenced by previous space-time measurements performed upon the same quantum particle? When a space-time measurement belongs to the class of space-time coordinates measurement, and when it belongs to the class of momentum or energy measurements (performed in an indirect manner using also space-time measurements)? Under which circumstances a measurement can be considered as an initial action (in this case its accuracy can be greater) or as a subsequent action (its accuracy having to be less than a certain value, according to Heisenberg relation)?. The time always appears in quantum mechanics, while two physical quantities cannot be measured *exactly* at the same moment of time.

So a space or time measurement performed at a certain time moment belongs to the class of subsequent indirect methods for measuring momentum or energy (having as a consequence a limited accuracy), or to the class of direct methods for measuring space or time (having a possible greater accuracy). A rigorous classification according to certain patterns should be made; see also [12], taking into consideration similarities in fundamental physics laws [13].

5.3. Different-scale system properties used for explaining logical aspects of pulse measurements

This problem suggests also a possible solution: if we continue our analysis of terms involved in measuring procedures, we can notice that both basic principles (light speed constance principle and uncertainty principle) use the term of measuring method. In an implicit manner, the terms *reference system* (for special relativity theory) and *measuring system* (for quantum theory) appear. Yet a measuring system implies the fact that it is not affected by the measuring procedure (otherwise, the physical quantity having to be measured would possess different values, depending on this interaction). So a first conclusion appears: the measuring system must be defined at a much larger scale than the body or the wave which interacts with it. The different scale system properties must be taken into consideration from the very beginning so as to put them into correspondence with

(i) the class of reference systems, which are not affected by interaction (where wave trains similar to wavelets can appear [14]);

(ii) the class of transient phenomena which undergo specific interactions (such transient phenomena can be represented as solitary waves, while estimations for the space coordinates for the source of the received wave-train based on space relations are not suitable for this purpose. As a further consequence, the constance light speed principle appears as a simple generation of another light wave when a received wave train arrives in the *material* medium of the observer reference system, and the uncertainty principle appears as a spread of a wave corresponding to a quantum particle by the measuring system, according to a kind of Fourier transformation performed on limited space and time intervals (the aperture and a certain working period). Thus logical aspects of the definitions of basic principles in physics (implying measurements of pulse parameters) can be explained in a rigorous manner.

6. Aspects connected with measurements on a set of pulses

6.1. Measurements on a set of pulses received on adjoining space-time intervals. Synchronization aspects

We will justify the previous considerations by presenting the case of measurements for sequence of pulses received on adjoining space-time intervals. As it is known, the special relativity theory considers that the Lorentz formulae describe the transformation of the space-time coordinates corresponding to an event when the inertial reference system is changed. These formulae are considered to be valid at any moment of time after a certain synchronization moment (the zero moment) irrespective to the measuring method used. However, there are some problems connected to the use of mechanical measurements on closed-loop trajectories. For example, let us consider that at the zero moment of time, in a medium with a gravitational field which can be neglected (the use of the Galileean form of the tensor g_{ik} being allowed) two observers are beginning a movement from the same point of space, in opposite directions, on circular trajectories having a very great radius of curvature. After a certain time interval, the observers are meeting again in the same point of space. For very great radii of curvature, the movements on very small time intervals can be considered as approximative inertial (as in the case of the transverse Doppler effect, where the time dilation phenomenon was noticed in the earth reference system which is approximative inertial on small time intervals). The Lorentz formulae can be applied on a small time interval $\Delta t(1)$ measured by one of the observers inside his reference system, and it results (using the Lorentz formula for time) that this interval corresponds to a time interval

$$\Delta t'(1) = \frac{\Delta t(1)}{\sqrt{1 - v(1)^2/c^2}} \tag{6.1}$$

in the reference system S_2 of the other observer, which moves with speed $v(1)$ as related to the reference system S_1 on this time interval. So the time dilation phenomenon appears. If each observer considers the end of this time interval ($\Delta t(1)$ or $\Delta t'(1)$) as a new zero moment (using a resynchronization procedure), the end of the second time interval $\Delta t(2)$ (with the new zero moment considered as origin) will correspond to a time moment

$$\Delta t'(2) = \frac{\Delta t(1)}{\sqrt{1 - v(2)^2/c^2}} \tag{6.2}$$

measured in the other reference system S_2 which moves with speed $v(2)$ as related to system S_1 on the time interval $\Delta t'(2)$ (with the new zero moment considered as origin). As related to the first zero moment (when the circular movement has started) the end of the second time interval appears at the time moment

$$t_2 = \Delta t(1) + \Delta t(2). \tag{6.3}$$

For the observers situated in reference system S_1, and at the time moment

$$t'(2) = \Delta t'(1) + \Delta t'(2) \frac{\Delta t(1)}{\sqrt{1 - v(1)^2/c^2}} + \frac{\Delta t(2)}{\sqrt{1 - v(2)^2/c^2}} \tag{6.4}$$

for the other observer.

Due to the fact that

$$\begin{aligned} \Delta t'(1) &> \Delta t(1), \\ \Delta t'(2) &> \Delta t(2), \end{aligned} \tag{6.5}$$

it results that

$$t'(2) = \Delta t'(1) + \Delta t'(2) > \Delta t(1) + \Delta t(2) = t(2) \tag{6.6}$$

and thus a global time dilation for the time interval $\Delta t(1) + \Delta t(2)$ appears. The procedure can continue, by considering the end of each time interval

$$\Delta t(1) + \Delta t(2) + \cdots + \Delta t(i) \tag{6.7}$$

as a new zero moment, and so it results that on all the circular movement period, a time moment

$$t(k) = \sum_{i=0}^{k} \Delta t(i) \tag{6.8}$$

(measured by the observer in reference system S_1) corresponds to a time moment

$$t'(k) = \sum_{i=0}^{k} \Delta t'(i) = \sum_{i=0}^{k} \frac{\Delta t(i)}{\sqrt{1 - v_i^2/c^2}} \tag{6.9}$$

(measured by the observer situated in reference system S_2) which implies

$$t'(k) > t(k). \tag{6.10}$$

By joining together all these time intervals $\Delta t(i)$ we obtain the period of the whole circular movement T. While the end of this movement is represented by the end of the time interval $\Delta t(N)$ in the reference system S_1, it results that T can be written under the form

$$T = t(N) = \sum_{i=0}^{N} \Delta t(i) \tag{6.11}$$

(considered in the reference system S_1) and it results also that this time moment (the end of the circular movement) corresponds to a time moment

$$T' = t'(N) = \sum_{i=0}^{N} \Delta t'(i) \tag{6.12}$$

measured in the reference system $S_@$. While

$$\Delta t'(i) = \frac{\Delta t(i)}{\sqrt{1 - v(i)^2/c^2}} > \Delta t(i), \tag{6.13}$$

it results

$$T' > T. \tag{6.14}$$

If the time is measured using the age of two twin children, it results that the twin in reference system S_2 is older than the other in reference system S_1, (having a less mechanical resistance of bones) and it can be destroyed by it after both observers stop their circular movements. However, the same analysis can be made starting from another set of small time intervals $\Delta_n t'(i)$ considered in the reference system S_2 which corresponds to a new set of time intervals $\Delta_n t(i)$ considered in the reference system S_2 (established using the same Lorentz relation) and finally it would result that the period of the circular movement T' measured in system S_2 corresponds to a period T greater than T' considered in reference system S_1. If the time is measured using the age of two twin children, it results that the twin in reference system S_1 is older than the other in reference system S_2, (having a less mechanical resistance of bones) and it can be destroyed by it after both observers stop their circular movements. But this result is in logic contradiction with the previous conclusion, because a man cannot destroy and in the same time be destroyed by another man [15].

As a first attempt of solving this contradiction, one can suppose that Lorentz formulae are valid only for electromagnetic phenomena (as in the case of the transversal Doppler effect) and not in case of mechanical phenomena. But such a classification is not a rigorous classification, being not suitable for formal logic. In next section, we will present a more rigorous classification of phenomena used in space-time measurements, which can be used for *gedanken* experiments using artificial intelligence based on formal logic.

6.2. Classification of space-time measurement methods based on memory of previous measurements

The logical contradiction presented in previous section appeared due to the fact that an element with internal memory has been used. The indication of this element has not been affected by the resynchronization procedure. In modern physics such an element with internal memory is connected with the corpuscular aspect of matter with a body. On the contrary, a measuring procedure based on an electromagnetic or optic wave train is a transient phenomenon. The synchronization of clocks is possible only after the wave-train arrives at the observer. Excepting a short time interval after the reception the received wave train does not exist inside the observer medium, so there is not any space area where a physical quantity which characterizes the wave to cumulate. That is the reason why a correct solution of the twins paradox must be based not on the association of electromagnetic (or optic) phenomena with the Lorentz formulae, but on the association of the Lorentz formulae with wave phenomena describing the propagation of a wave inside the observers reference systems. The wave class is more general than the class of electromagnetic and optic waves (we can mention the wave associated with particles in quantum mechanics). Besides, in the most general case, the interaction between two reference systems appears under the form of a field, not under the form of a material body. Moreover, this aspect implies an intuitive interpretation for the dependence of the mass of a body inside a reference system.

Using the formal logic, all we have shown can be presented in a rigorous manner.

(a) We define the notion of "propagation" phenomenon in two inertial reference systems (the system where the event takes place and the system where a signal generated by the event is noticed) as a phenomenon having a finite existence inside the reference system, the number of intervals being finite.

(b) We define the notion of corpuscle inside a certain reference system as a phenomenon which can possess an unlimited evolution in time and space inside the reference system; it can be also said that the phenomenon has its own existence, it exists by itself.

(c) We define the emission of a wave-train U_e in a reference system and its transformation in another train when it interacts with the observers medium

Definition 6.1. There exist an area S_{0e} and a time interval T_{0e} in the reference system where the emission takes place so that

$$F_{ue}(S_{0e}, T_{0e}) \neq 0, \quad F_{ue}(S_{0e}, t) = 0 \quad \text{for } t \notin T_{0e}. \tag{6.15}$$

There exist a space area S_{0r} and a time interval T_{0r} in the observer reference system, and a relation Tr so that

$$\begin{gathered} F_{ur}(S_{0r}, T_{0r}) = \mathrm{Tr}[F_{ue}(S_{0e}, T_{0e})], \\ F_{ur}(S_{0r}, T_{0r}) \neq 0, \quad F_{ur}(S_{0r}, t) = 0 \quad \text{for } t \notin T_{0r}. \end{gathered} \tag{6.16}$$

(d) We define the transformation of a sequence of received pulses $\Sigma_k Ue_k$ in a sequence $\Sigma_k Ur_k$, $k = 1 \cdots n$ after interaction with the observers reference system, by considering that each pulse (wave-train) is transformed in an independent manner by the material medium of the observer reference system, according to its specific Lorentz transformation

Definition 6.2. Consider

$$\begin{gathered} Ur_k = L_k[Ue]_k, \\ \Sigma_k Ue_k = \Sigma_k Ur_k, \end{gathered} \tag{6.17}$$

where L_k represents the Lorentz transformation performed upon the Ue_k wave by the system with the interaction moment of this wave with the material medium of the observer considered as zero moment of time (synchronization moment) for the Lorentz transformation L_k.

(e) We define the interaction between a sequence of pulses and the material body of the observer reference system (a corpuscle) as an interaction function Int between the material medium and each transformed pulse Ur_k corresponding to a received pulse Ue_k, the mass m of the body measuring the influence of the received wave-train Ue_k upon the body.

Definition 6.3. Consider

$$\frac{1}{m} = \operatorname{Int}[Ur_k] = \operatorname{Int}[L_k(Ue)_k]. \tag{6.18}$$

When Lorentz transformation L_k does not generate a pulse Ur_k (e.g., when the relative speed between the material body and the wave is equal to c, the speed of light in vacuum), the mass m is equal to ∞, which means that no interaction due to the received pulse Ue_k exists (an idea which connects the notion on infinite mass with the absence of interaction). So $m = \infty$ for a body inside a reference system S shows that we cannot act upon the material body using wave pulses emitted in system S; however, changes in the movement of the body (considered in system S) due to other external forces seem to be allowed.

By interaction with a certain material medium, each pulse is transformed according to Lorentz formulae, and the modified parameters of each pulse must replace the previous informations in the memory cells.

7. Associating a certain wave train to Lorentz transformation

7.1. The necessity for associating a wave function to the Lorentz transformation

The Lorentz transformation is usually represented as a matrix L which acts upon a quadridimensional column vector r having the components $r_1 = x$, $r_2 = y$, $r_3 = z$, $r_4 =$ ict, resulting in another quadridimensional vector r^* having the components $r_1' = x'$, $r_2' = y'$, $r_3' = z'$, $r_4' =$ ict$'$, where x, y, z, t are the space-time coordinates corresponding to a certain event in an inertial reference system S, and x', y', z', t' are the space-time coordinates corresponding to the same event measured in an inertial reference system S' which moves with velocity v (a vector) against the system S. This means

$$r' = L(v)r. \tag{7.1}$$

All time moments are considered after a synchronization moment (when the clock indications in the reference systems are set to zero). The velocity v defines the matrix L, and the result is considered not to depend on the measuring method used. But let us consider that the velocity v has two components v_x and v_y oriented along the Ox axis (for v_x) and along the Oy axis (for v_y) and let us consider also that the event taking place in the reference system S is first observed in a reference system S_1 which moves with velocity v_x as against the system S :

a set of space-time coordinates (x_1, y_1, z_1, t_1) will be established for the event. Then the event having the space-time coordinates (x_1, y_1, z_1, t_1) in system S_1 is observed in the reference system S' which moves with velocity v_y (the projection of v along the Oy axis) against the reference system S (this relative speed being measured in system S). That corresponds to a relative speed

$$v_y(c) = \frac{v_y}{\sqrt{1 - v_x^2/c^2}} \tag{7.2}$$

between the systems S and S' (due to the kinematics law of addition of speeds in special relativity theory). Thus will result in the cuadridimensional vector r' (having the components x', y', z', ict$'$), measured in system S', under the form

$$r' = L(v_y(c))L(v_x)r. \tag{7.3}$$

But we can also consider that the event having the space-time coordinates x, y, z, t in system S is first observed in a reference system S_2 which moves with velocity v_y (the projection of velocity v along the Oy axis) as against system S; a set of space-time coordinates will be established for the event. Then this event having the space-time coordinates x_2, y_2, z_2, t_2 in system S_2 is observed in the reference system S' which moves with velocity v_x (the projection of velocity v along the Ox axis) against the reference system S_2, the velocity vx being measured in the reference system S. That corresponds to a relative speed

$$v_x(c) = \frac{v_x}{\sqrt{1 - v_y^2/c^2}} \tag{7.4}$$

between the systems S' and S_2 (due to the same kinematics law of addition of speeds in special relativity). Thus will result in the space-time coordinates x', y', z', t' measured in system S' under the form

$$r' = L(v_x(c))L(v_y)r. \tag{7.5}$$

Using the explicit form of Lorentz transformation for the case when the relative speed has the direction of one of the axes of coordinates, it can be easy shown that

$$L(v_y(c))L(v_x)r \neq L(v_x(c))L(v_y)r. \tag{7.6}$$

This shows that the coordinates measured for the event in S' reference system depends on the succession of transformations. This aspect is similar to the noncommutative properties of operators in quantum theory [16]. It implies that in the case of special relativity we must define a vector of state (a wave-function) upon which the Lorentz transformation acts. Thus the Lorentz transformation can be considered as a physical transformation which modifies a certain wave function inside a reference system. Taking into account the fact that usually we receive information under the form of electromagnetic (or light) wave trains (the emission of these wave trains corresponding to the event) and taking also into account the fact that the time-dilation phenomenon (a consequence of Lorentz transformation) was first time observed for light wave trains (the transverse Doppler effect) it results that in the most general case this wave function must be associated to the wave-function of the received light wave train. As a consequence of the previous statement, it results that a Lorentz transformation L must be always put in correspondence with a pair (S, φ), S representing a certain material reference system which acts upon a wave train having the state-vector φ. So the Lorentz transformation must be written under the form $L_S(\varphi)$; in the most general case L is the Lorentz matrix and φ is a vector or a higher-order tensor which describes the field. For an electromagnetic wave, the field can be described using the cuadridimensional vector A. The action of the matrix L_S consists in a general transformation

$$\varphi(x, y, z, t) \longrightarrow \varphi'(x', y', z', t') = L_S\varphi(x, y, z, t), \tag{7.7}$$

where the values of φ are modified according to the transformation rules of vectors and tensors (e.g., $A' = LA$ for an electromagnetic wave described by the cuadrivector A) and in the change of the space-time coordinates (x, y, z, t) into (x', y', z', t') according to the formula

$$r' = L_S r, \tag{7.8}$$

r representing the cuadridimensional vector of coordinates. We have to point the fact that in all these formulae $\varphi(x, y, z, t)$ represents the value φ would have possessed in the absence of the interaction with the observer material medium; the space-time origin must be considered in the point of space and at the moment of time where the wave first time interacts with the observer material medium (in a similar way with the aspects in quantum mechanics, where all transformations are acting after the interaction with the measuring system). This interpretation can solve the contradictions appearing in case of movements on closed-loop trajectories (the twins paradox) in a very simple manner. The Lorentz transformation being a transformation which acts upon a certain wave train (a light wave train, in the most general case), it has no consequences upon the age of two observers moving on closed-loop trajectories. So no contradiction can appear when the two observers are meeting again.

7.2. Possibilities of using the principle of least action in connection with the wave-train interpretation

We begin by writing the propagation equation for an electromagnetic wave inside an observer material medium under the form $dx^2 + dy^2 + dz^2 = c^2 dt_2$ (c representing the light speed). It results that $c^2dt^2 - dx^2 - dy^2 - dz^2 = 0$ for all points inside the material medium where the wave has arrived. But

$$c^2dt^2 - dx^2 - dy^2 - dz^2 = ds^2, \tag{7.9}$$

where ds is the cuadridimensional space-time interval. The propagation equation of the optical wave can be written as $ds = 0$, and so it results that the trajectory of the wave inside the material medium between two points a and b is determined by the equation

$$\int_a^b ds = \Delta s = 0. \tag{7.10}$$

By the other hand, for mechanical phenomena the quantity determining the trajectory of a material body inside a reference system is the action S. Under a relativistic form, it can be written as $S = -mc\int_a^b ds$, m representing the mass of the body, and a, b the space-time coordinates for two points situated along the "universe line" on which the body moves. The principle of least action can be written as $\delta S = -mc\delta\int_a^b ds = 0$.

While $\delta S = \sum_i mcu_i\delta x_i$, where $u_i = v_i/\sqrt{1 - v^2/c^2}$ for $i = 1, 2, 3$ and $u_4 = ic/\sqrt{1 - v^2/c^2}$, it results finally that $\sum_i p_i^2 = -m^2c^2$, p_i being the cuadrivector $\partial S/\partial x_i$ (the momentum). For a free particle, $p_i = mu_i$. It can be noticed that the infinite small cuadridimensional interval ds is used both for describing the propagation of an electromagnetic wave and the movement of a body inside a reference system. While is it related to the action S, this result is easy to be understood (the principle of least action being a basic principle in nature). The next step consists in pointing the fact that the previous integral $\Delta s = 0$ (determining the trajectory of the optical wave train inside the material medium) is based on the supposition that both points a, b belong to the material medium (otherwise, the velocity of the wave may differ, depending on the dielectric and magnetic constants of the material). So the equation can be directly used in measurement procedures (for establishing trajectory or other properties of the wave only for the time interval when the optical wave train exists in that material medium [17]). If an observer has to analyze a wave train emitted in another material reference system, he must use the invariance property of the cuadriinterval: $ds = ds'$, where ds represents the cuadriinterval between two close events in a certain inertial reference system and ds' represents the cuadriinterval between the same two events measured in another reference system. While $ds = ds(dx, dy, dz, dt)$ is determined inside the observer reference system and $ds' = ds'(dx', dy', dz', dt')$ corresponds to the reference system where the wave has been emitted, it results that the cuadridimensional interval ds moves into the cuadridimensional interval ds' by a function

$$ds(dx, dy, dz, dt) \Longrightarrow L \Longrightarrow ds'(dx', dy', dz', dt'), \tag{7.11}$$

where the arguments of ds are transformed by the Lorentz relations

$$dx' = \frac{dx + vdt}{\sqrt{1 - v^2/c^2}}, \quad dy' = dy, \quad dz' = dz, \quad dt' = \frac{dt + vdx/c^2}{\sqrt{1 - v^2/c^2}} \tag{7.12}$$

for v parallel to Ox (all the space and time intervals dx, dy, dz, dt being considered inside the observer material medium after the emitted optical wave train arrives), and $ds = ds'$. The above relation can be considered as presenting a transformation of the received wave train (with x, y, z, t coordinates) into a "supposed" wave train corresponding to the case when the wave train would not have entered inside the observer material medium. For determining the real trajectory of the wave before interaction the observer must extend the trajectory of the received wave train (having coordinates x', y', z', t' in the past and outside the observer material medium, using the relation

$$\int_a^b ds' = \Delta s' = 0. \tag{7.13}$$

7.3. Non-Markov aspects of pulse transformation

We have also to emphasize the non-Markov aspect of Lorentz transformation which acts upon a received wave train when this interacts with the observer material medium. At the initial moment of time (the zero moment) we can consider that new values for wave quantities are generated as a result of the Lorentz matrix action upon the received values (cuadrivectors or cuadritensors). This represents a Markov transformation (using some physical quantities defined at a certain moment of time $t = 0$, we can obtain the result of that transformation at a time moment $t + dt = 0 + dt$).

Yet if we analyze the wave train transformation at a subsequent moment of time (after the zero moment when the wave was received) we can notice that the physical quantities corresponding to cuadrivectors and cuadritensors are not just modified (by the action of Lorentz matrix) but are also translated at a different time moment (according to Lorentz formulae for transforming space-time coordinates). This implies that the physical quantities corresponding to the transformed wave train (defined in the observer material reference system) depend on the physical quantities corresponding to the unchanged wave train (supposed situation) at a previous time moment. Not being possible to use values of certain quantities at a time moment t for obtaining the values of that physical quantities at a time moment $t + dt$ for $t > 0$, it results that the Lorentz transformation of a received wave train (an electromagnetic or optic pulse or an associated wave corresponding to a particle) is a non-Markov transformation. In future studies, this aspects should be studied using aspects connected to time series inside a material medium [18].

8. Conclusions

This study has shown that certain intuitive problems connected with measurements of sequences of pulses on closed-loop trajectories in special relativity and noncommutative properties of operators in quantum physics imply a more rigorous definition of measurement method and of the interaction phenomena (classified according to a possible memory of previous measurements), so as to avoid logical contradictions due to a possible resynchronization. It is also shown that the use of the least action principle requires a specific space-time interval available for a space-time measurement in an implicit form. Due to this, it results in a certain distinction between the set of existing space-time intervals (which can be defined on unlimited space-time intervals) and the set of measured space-time intervals (established using measuring methods based on waves and always defined on limited space-time intervals).

References

[1] E. Bakhoum, “Fundamental disagreement of wave mechanics with relativity,” *Physics Essays*, vol. 15, no. 1, pp. 87–100, 2002.

[2] P. A. M. Dirac, *The Principles of Quantum Mechanics*, Oxford University Press, Oxford, UK, 1958.

[3] R. Feynman, *The Feynman Lectures on Physics*, vol. 2, Addison Wesley, Reading, Mass, USA, 1964.

[4] Y. Aharonov and D. Bohm, “Significance of electromagnetic potentials in the quantum theory,” *Physical Review*, vol. 115, no. 3, pp. 485–491, 1959.

[5] Y. Imry and R. A. Webb, “Quantum interference and the Aharonov-Bohm effect,” *Scientific American*, vol. 260, no. 4, pp. 56–62, 1989.

[6] A. P. French and E. F. Taylor, *An Introduction to Quantum Physics*, Norton Publications, New York, NY, USA, 1978.

[7] L. de Broglie, *New Perspectives in Physics*, Basic Books, New York, NY, USA, 1962.

[8] L. de Broglie, *The Current Interpretation of Wave Mechanics: A Critical Study*, Elsevier, Amsterdam, The Netherlands, 1964.

[9] D. Grifiths, *Introduction to Elementary Particles*, John Wiley & Sons, New York, NY, USA, 1987.

[10] C. Morarescu, “Inner potential of generating pulses as a consequence of recurrent principles and specific computing architecture,” in *Computational Science and Its Applications*, vol. 3980 of *Lecture Notes in Computer Science*, pp. 814–820, Springer, Berlin, Germany, 2006.

[11] C. Toma, “A connection between special relativity and quantum theory based on non-commutative properties and system—wave interaction,” *Balkan Physics Letters*, vol. 5, pp. 2509–2513, 1997.

[12] M. Takeda, Sh. Inenaga, and H. Bannai, “Discovering most classificatory patterns for very expressive pattern classes,” in *Discovery Science*, vol. 2843 of *Lecture Notes in Computer Science*, pp. 486–493, Springer, Berlin, Germany, 2003.

[13] C. Toma, “The advantages of presenting special relativity using modern concepts,” *Balkan Physics Letters*, vol. 5, pp. 2334–2337, 1997.

[14] C. Cattani, “Harmonic wavelets towards the solution of nonlinear PDE,” *Computers & Mathematics with Applications*, vol. 50, no. 8-9, pp. 1191–1210, 2005.

[15] M. Simeonidis, S. Pusca, G. Toma, A. Toma, and T. Toma, “Definition of wave-corpuscle interaction suitable for simulating sequences of physical pulses,” in *Computational Science and Its Applications*, vol. 3482 of *Lecture Notes in Computer Science*, pp. 569–575, Springer, Berlin, Germany, 2005.

[16] P. Sterian and C. Toma, “Methods for presenting key concepts in physics for MS students by Photon-MD program,” *Bulgarian Journal of Physics*, vol. 27, no. 4, pp. 27–30, 2000.

[17] C. Toma, “The use of the cuadridimensional interval—the main possibility for improving the Lorentz formulae interpretation,” in *Proceedings of the European Conference on Iteration Theory (ECIT ’97)*, vol. 2, p. 202, Pitesti, Romania, November 1997.

[18] M. Olteanu, V.-P. Paun, and M. Tanase, “The analysis of the time series associated to sem microfractographies of Zircaloy-4,” *Revista de Chimie*, vol. 56, no. 7, pp. 781–784, 2005.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2009, Article ID 359481, 14 pages
doi:10.1155/2009/359481

Research Article

Vanishing Waves on Closed Intervals and Propagating Short-Range Phenomena

Ghiocel Toma[1] and Flavia Doboga[2]

[1] *Faculty of Applied Sciences, Politechnica University, 061071 Bucharest, Romania*
[2] *Modeling and Simulation Department, ITT Industries, Washington, DC 20024, USA*

Correspondence should be addressed to Flavia Doboga, flaviactrdoboga@yahoo.com

Received 29 May 2008; Accepted 24 June 2008

Recommended by Carlo Cattani

This study presents mathematical aspects of wave equation considered on closed space intervals. It is shown that a solution of this equation can be represented by a certain superposition of traveling waves with null values for the amplitude and for the time derivatives of the resulting wave in the endpoints of this interval. Supplementary aspects connected with the possible existence of initial conditions for a secondorder differential system describing the amplitude of these localized oscillations are also studied, and requirements necessary for establishing a certain propagation direction for the wave (rejecting the possibility of reverse radiation) are also presented. Then it is shown that these aspects can be extended to a set of adjacent closed space intervals, by considering that a certain traveling wave propagating from an endpoint to the other can be defined on each space interval and a specific mathematical law (which can be approximated by a differential equation) describes the amplitude of these localized traveling waves as related to the space coordinates corresponding to the middle point of the interval. Using specific differential equations, it is shown that the existence of such propagating law for the amplitude of localized oscillations can generate periodical patterns and can explain fracture phenomena inside materials as well.

1. Introduction

Test-functions(whichdifferto zero only on a limited interval and have continuous derivatives of any order on the whole real axis) are widely used in the mathematical theory of distributions and in Fourier analysis of wavelets. Yet such test-functions, similar to the Dirac functions, cannot be generated by a differential equation. The existence of such an equation of evolution, beginning to act at an initial moment of time, would imply the necessity for a derivative of certain order to make a jump at this initial moment of time from the zero value to a nonzero value. But this aspect is in contradiction with the property of test-functions to have continuous derivatives of any order on the whole real axis, represented in this case by the time axis. So it results that an ideal test-function cannot be generated by a differential equation (see also [1]); the analysis has to be restricted at possibilities of generating practical test-functions (functions similar to test-functions, but having a finite number of continuous derivatives on the whole real axis) useful for wavelets analysis. Due to the exact form of the derivatives of test-functions, we cannot apply derivative free algorithms [2] or algorithms which can change in time [3]. Starting from the exact mathematical expressions of a certain test-function and of its derivatives, we must use specific differential equations for generating such practical test-functions.

Thisaspect is connected with causal aspects of generating apparently acausal pulses as solutions of the wave equation, presented in [4]. Such test-functions, considered at the macroscopic scale (that does not mean Dirac-functions), can represent solutions for certain equations in mathematical physics (an example being the wave-equation). The main consequence of this aspect consists in the possibility of certain pulses to appear as solutions of the wave-equation under initial null conditions for the function and for all its derivatives and without any free-term (a source-term) to exist. In order to prove the possibility of appearing acausal pulses as solutions of the wave-equation (not determined by the initial conditions or by some external forces) we begin by writing the wave-equation

$$\frac{\partial^2 \phi}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 \phi}{\partial t^2} = 0 \tag{1.1}$$

for a free string defined on the length interval $(0, l)$ (an open set), where ϕ represents the amplitude of the string oscillations and v represents the velocity of the waves inside the string medium. At the initial moment of time (the zero moment) the amplitude ϕ together with all its derivatives of first and second orders is equal to zero. From the mathematical theory of the wave-equation, we know that any solutionof this equation must be a superposition of a direct wave and of a reverse wave. For the beginning, we will restrict our analysis at direct waves by considering a supposed extension of the string on the whole Ox axis, ϕ being defined by the function

$$\phi(\tau) = \begin{cases} \exp\left(\dfrac{1}{(x - vt + 1)^2 - 1}\right) & \text{for } |x - vt + 1| < 1, \\ 0 & \text{for } |x - vt + 1| \geq 1, \end{cases} \tag{1.2}$$

where $t \geq 0$. This function for the extended string satisfies the wave-equation (being a function of $x - vt$, a direct wave). It is a continuous function, having continuous partial derivatives of any order for $x \in (-\infty, \infty)$ and for $t \geq 0$. For $x \in (0, l)$ (the real string) the amplitude ϕ and all its derivatives are equal to zero at the zero moment of time, as required by the initial null conditions for the real string (nonzero values appearing only for $x \in (-2, 0)$ for $t = 0$, while on this interval $|x - vt + 1| = |x + 1| < 1$). We can notice that for $t = 0$ the amplitude ϕ and its partial derivatives differ to zero only on a finite space interval, this being a property of the functions defined on a compact set (test-functions). But the argument of the exponential function is $x - vt$; this implies that the positive amplitude existing on the length interval $(-2, 0)$ at the zero moment of time will move along the Ox axis in the direction $x = +\infty$. So at some time moments t_k after the zero moment, a nonzero amplitude ϕ will appear inside the string, propagating from one edge to the other. It can be noticed that the pulse passes through the real string and at a certain time moment t_{fin} (when the pulse existing at the zero moment of time on the length interval $(-2, 0)$ has moved into the length interval $(l, l + 2)$) its action upon the real string ceases. We must point the fact that the limit points $x = 0$ and $x = l$ are not considered to belong to the string; but this is in accordance with the rigorous definition of derivatives (for this limit points cannot be defined as derivatives related to any direction around them).

This point of space (the limit of the open space interval considered) is very important for our analysis, while we will extend the study to closed space intervals. Considering small space intervals around the points of space where the sources of the generated field are situated (e.g., the case of electrical charges generating the electromagnetic field), it will be shown that causal aspects require the logical existence of a certain causal chain for transmitting interaction from one point of space to another, which can be represented by mathematical functions which vanish (its amplitude and all its derivatives) in certain points of space. From this point of space, an informational connection for transmitting the wave further could be considered (instead of a transmission based on certain derivatives of the wave). Thus a kind of granular approach for propagation along a certain axis can be considered suitable for application in quantum theory. As an important consequence, some directions of propagation for the generated wave will appear and the possibility of reverse radiation will be rejected. Moreover, specific applications for other propagating phenomena involving the generation of some spatial periodical patterns or an increasing amplitude of oscillations along a certain spatial axis can be also analyzed by this mathematical model.

2. Utility of test-functions in mathematical physics for half-closed space intervals

If we extend our analysis to half-closed intervals by adding one endpoint of the space interval to the previously studied open intervals (e.g., by adding the point $x = 0$ to the open interval $(0, l)$), we should take into account the fact that a complete mathematical analysis usually implies the use of a certain function $f(t)$ defined at the limit of the working space interval (the point of space $x = 0$, in the previous example). Some other supplementary functions can be met in mathematical physics.

The use of such supplementary functions defined on the limit of the half-closed interval could appear as a possible explanation for the problem of generating acausal pulses as solutions of the wave equation on bounded open intervals. The acausal pulse presented in the previous paragraph (similar to wavelets) traveling along the Ox axis requires a certain nonzero function of time $f_0(t)$ for the amplitude of the pulse for the limit of the interval $x = 0$. It could be argued that the complete mathematical problem of generating acausal pulses for null initial conditions on this interval and for null function $f_0(t)$ corresponding to function ϕ (the pulse amplitude) at this endpoint of the interval ($x = 0$, resp.) would reject the possibility of appearing the acausal pulse presented in the previous paragraph. The acausal pulse ϕ previously presented implies nonzero values for f_0 at certain time moments, which represents a contradiction with the requirement for this function f_0 to present null values at any time moment. By an intuitive approach, null external sources would imply null values for function f_0 and (as a consequence) null values for the pulse amplitude ϕ.

Yet it can be easily shown that the problem of generating acausal pulses on half-closed intervals cannot be rejected by using supplementary requirements for certain functions $f(t)$ defined at one limit of such bounded space intervals. Let us simply suppose that instead of function

$$\phi(\tau) = \begin{cases} \exp\left(\dfrac{1}{(x - vt + 1)^2 - 1}\right) & \text{for } |x - vt + 1| < 1, \\ 0 & \text{for } |x - vt + 1| \geq 1 \end{cases} \tag{2.1}$$

presented in the previous paragraph we must take into consideration two functions ϕ_0 and ϕ_l defined as

$$\phi_0(\tau) = \begin{cases} \exp\left(\dfrac{1}{(x - vt + m)^2 - 1}\right) & \text{for } |x - vt + m| < 1, \\ 0 & \text{for } |x - vt + m| \geq 1, \end{cases}$$
$$\phi_l(\tau) = \begin{cases} -\exp\left(\dfrac{1}{(x + vt - m)^2 - 1}\right) & \text{for } |x - vt + m| < 1, \\ 0 & \text{for } |x + vt - m| \geq 1, \end{cases} \tag{2.2}$$

with m selected as $m > 0$, $m - 1 > l$ (so as both functions ϕ_0 and ϕ_l to have nonzero values outside the real string and asymmetrical as related to the point of space $x = 0$. While function ϕ_0 corresponds to a direct wave (its argument being $(x - vt)$) and ϕ_l corresponds to a reverse wave (its argument being $(x + vt)$) it results that both functions ϕ_0 and ϕ_l arrive at the same space origin $x = 0$, the sum of these two external pulses being null all the time (functions ϕ_0 and ϕ_l being asymmetrical, $\phi_0 = -\phi_l$) at any moment of time. So by requiring that $\phi(t) = 0$ for $x = 0$ (the left limit of a half-closed interval $[0, l)$) we cannot reject the mathematical possibility of the appearance of an acausal pulse on a half-closed interval.

A possible mathematical explanation for this aspect consists in the fact that we have used a reverse wave (an acausal pulse) propagating from $x = \infty$ toward $x = -\infty$, which is first received at the right limit $x = l$ of the half-closed interval $[0, l)$ before arriving at the point of space $x = 0$. It can be argued that in case of a closed space interval $[0, l]$, we should consider the complete mathematical problem, consisting of two functions $f_0(t)$, $f_l(t)$ corresponding to both limits of the working space intervals (the points of space $x = 0$ and $x = l$). But in fact the wave equation corresponds to a physical model valid in the three-dimensional space, under the form

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2 \phi}{\partial t^2} = 0 \tag{2.3}$$

and the one-dimensional model previously used is just an approximation. Moreover, the source of the field is considered at a microscopic scale (e.g., quantum particles like electrons for the case of the electromagnetic field) and the emitted field for such elementary particles presents a spherical symmetry. Transforming the previous equation in polar coordinates and supposing that the function ϕ depends only on r (the distance from the source of the field to the point of space where this emitted field is received), it results that

$$\frac{\partial^2 U}{\partial r^2} - \frac{1}{v^2}\frac{\partial^2 U}{\partial t^2} = 0, \tag{2.4}$$

where

$$U = r\varphi. \tag{2.5}$$

An analysis of the field emitted from the point of space $r = 0$ (the source) toward a point of space $r = r_0$ (where the field is received) should be performed on the space interval $(0, r]$ (a half-closed interval); the point of space $r = 0$ cannot be included in the working interval as long as the solution $\phi(r)$ for the field is obtained by dividing the solution $U(r)$ of the previous equation (in spherical coordinates) through r (the denominator of the solution ϕ being zero, some supplementary aspects connected to the limit of functions should be added, but still without considering a function of time as condition for the space origin). This can be put in correspondence with the previously presented case of an acausal pulse defined on $[0, l)$ if we consider that (as a rule) (a) the endpoint where the function $\phi(t)$ is not defined represents the source of the field (a round bracket being added, while it cannot be considered as part of the working interval) and (b) the endpoint where the function φ vanishes represents a point of space where the propagating phenomenon is recreated (by reflection or by interaction with different particles, for the case of optical waves), a square bracket being added. The endpoint represented by square bracket (where the wave vanishes) can be considered as a source for the field propagating in a next space interval after an interaction, and so on.

Thus an asymmetry in the required methods for analyzing phenomena appears. Moreover, for the appearance of a certain direction for the transmission of interaction (from one space interval to another), it results that the possibility of retroradiation (a reverse wave generated by points of space where a direct wave has arrived) should be rejected (a memory of previous phenomena is determining the direction of propagation).

3. Applications for closed space intervals: applications in quantum physics

The pulse presented in the previous paragraph is in fact a traveling wave propagating from $x = \infty$ toward $x = 0$ and back which vanishes at the point of space $x = 0$ due to a kind of reflection. Yet we can extend our analysis by considering a subsequent reflection of this pulse at the limit point $x = l$ and so on. Thus a resulting traveling wave can be considered inside the *closed* space interval $[0, l]$ with null values at the endpoints $x = 0$, $x = l$ at any time moment *after* the first reflection.

At first sight, this localized oscillation is not useful for our mathematical analysis of acausal pulses. It does not correspond to initial null conditions on the closed bounded space interval $[0, l]$ and to null time functions defined at the endpoints $x = 0, x = l$ (while the traveling wave should already exist inside this interval when null conditions for the endpoints at any subsequent time moment are added). Yet we must take into consideration the fact that in quantum physics the operators corresponding to creation and annihilation of particles are obtained (in a heuristic manner) starting from an analysis of electromagnetic field performed on bounded space intervals and extended to unbounded intervals by simply replacing the space limits for a set of such intervals with infinite values [5]. However, the previously mentioned analysis on bounded intervals makes use of stationary waves which cannot be taken into consideration when a space limit equals $\pm\infty$ (no reflection can appear). This logical contradiction can be avoided if any extended space interval is considered as a sum of adjacent small space intervals with specific localized oscillations defined on each of them.

Supposing that a localized oscillation is generated on a certain limited space interval by an external force or by a received wave-train, we can consider that subsequent oscillations are generated on adjacent space intervals (as in the case of spherical waves) due to a kind of informational connection existing on the boundaries of these intervals. A mathematical connection described by wave-equation cannot be taken into consideration any more, and thus the previous model of causal chain corresponding to a sequence: *changes in the value of partial derivatives as related to space coordinates imply changes in the partial derivatives of the amplitude as* *related to time, which further imply changes in the value of the function,* should be replaced by a step-by-step transmission of interaction starting from an initial half-closed interval (e.g., its open limit corresponding to the source of the field) to adjacent space intervals. This corresponds to a granular aspect of space suitable for applications in quantum physics, where the generation and annihilation of quantum particles should be considered on limited space-time intervals (asymmetrical pulses could be also used [6]). A specific physical quantity (corresponding to the amplitude of localized oscillations) is transmitted from one space interval to another, according to a certain mathematical law.

4. Dynamical spatial generation of structural patterns

We will continue the study by presenting properties of spatial linear systems described by a certain physical quantity generated by a differential equation. This quantity can be represented by internal electric or magnetic field inside the material or by similar physical quantities, and corresponds to the amplitude of localized oscillations previously mentioned. A specific mathematical law which can be approximated by a differential equation generates this quantity considering as input the spatial alternating variations of a certain internal parameter. As a consequence, specific spatial linear variations of the corresponding physical quantity appear. In case of very short-range variations of this internal parameter, systems described by a differential equation able to generate a practical test-function [1] exhibit an output which appears to an external observer under the form of two distinct envelopes. These can be considered as two distinct structural patterns located in the same material along a certain linear axis. This aspect differs from the oscillations of unstable type second-order systems studied using difference equations [7] or advanced differential equations [8], and they differ also from the previous studies of the same author [9] where the frequency response of such systems to alternating inputs was studied (in conjunction with the ergodic hypothesis). For our purpose, we will use the function

$$\varphi(x) = \begin{cases} \exp\left(\dfrac{1}{x^2-1}\right) & \text{if } x \in (-1,1), \\ 0 & \text{otherwise,} \end{cases} \tag{4.1}$$

which is a test-function on $[-1,1]$. For a small value of the numerator of the exponent, a rectangular shape of the output is obtained. An example is the case of the function

$$\varphi(x) = \begin{cases} \exp\left(\dfrac{0.1}{x^2-1}\right) & \text{if } x \in (-1,1), \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

Using the expression of $\varphi(x)$ and of its derivatives of first and second orders, a differential equation which admits as solution the function φ corresponding to a certain physical quantity can be obtained. However, a test-function cannot be the solution of a differential equation. Such an equation of evolution implies a jump at the initial space point for a derivative of certain order, and test-function must possess continuous derivatives of any order on the whole real axis. So it results that a differential equation which admits a test-function φ as solution can generate only a practical test-function f similar to φ, but having a finite number of continuous derivatives on the real Ox axis. In order to do this, we must add initial conditions for the function f (generated by the differential equation) and for some of its derivatives $f^{(1)}$, and/or $f^{(2)}$ and so on equal to the values of the test-function φ and of some of its derivatives $\varphi^{(1)}$, and/or $\varphi^{(2)}$ and so on at an initial space point x_{in} very close to the beginning of the working spatial interval. This can be written under the form

$$f_{x_{\text{in}}} = \varphi_{x_{\text{in}}}, \quad f^{(1)}_{x_{\text{in}}} = \varphi^{(1)}_{x_{\text{in}}}, \quad \text{and/or} \quad f^{(2)}_{x_{\text{in}}} = \varphi^{(2)}_{x_{\text{in}}}, \quad \text{and so on.} \tag{4.3}$$

If we want to generate spatial practical test-functions f which are symmetrical as related to the middle of the working spatial interval, we can choose as space origin for the Ox axis the middle of this interval, and so it results that the function f should be invariant under the transformation

$$x \longrightarrow -x. \tag{4.4}$$

Functions invariant under this transformation can be written in the form $f(x^2)$ (similar to aspects presented in [1]) and so the form of a general second-order differential equation generating this kind of functions should be

$$a_2(x^2)\frac{d^2 f}{d(x^2)^2} + a_1(x^2)\frac{df}{dx^2} + a_0(x^2)f = 0. \tag{4.5}$$

However, for studying the generation of structural patterns on such a working interval, we must add a free term corresponding to the cause for the variations of the external observable physical quantity. Thus, a model for generating a practical test-function using as input the internal parameter $u = u(x)$, $x \in [-1, 1]$, is

$$a_2(x^2)\frac{d^2 f}{d(x^2)^2} + a_1(x^2)\frac{df}{dx^2} + a_0(x^2)f = u \tag{4.6}$$

subject to

$$\lim_{x\to\pm 1} f^k(x) = 0 \quad \text{for } k = 0, 1, \ldots, n, \tag{4.7}$$

which are the boundary conditions of a practical test-function. For u represented by alternating functions, we should notice periodical variations of the external observable physical quantity f.

According to the previous considerations for the form of a differential equation invariant at the transformation

$$x \longrightarrow -x, \tag{4.8}$$

a first-order system can be written under the form

$$\frac{df}{d(x^2)} = f + u \tag{4.9}$$

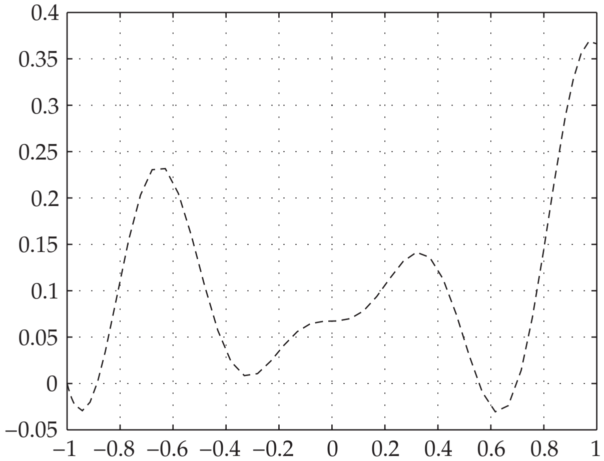


Figure 1: f versus distance for first-order system, input $u = \sin(10x)$.

which converts to

$$\frac{df}{dx} = 2xf + 2xu \tag{4.10}$$

representing a first-order dynamical system. For a periodical input (corresponding to the internal parameter) $u = \sin 10x$, numerical simulations performed using Runge-Kutta functions in MATLAB present an output of an irregular shape (Figure 1) not suitable for joining together the outputs for a set of adjoining linear intervals (the value of f at the end of the interval differs in a significant manner to the value of f at the beginning of the interval). A better form for the physical quantity f is obtained for variations of the internal parameter described by the equation $u = \cos 10x$. In this case, the output is symmetrical as related to the middle of the interval (as can be noticed in Figure 2) and the results obtained on each interval can be joined together on the whole linear spatial axis, without any discontinuities to appear. The resulting output would be represented by a sum of two great spatial oscillations (one at the end of an interval and another one at the beginning of the next interval) and two small spatial oscillations (around the middle of the next interval).

Similar results are obtained for an undamped dynamical system first order, represented by

$$\frac{df}{d(x^2)} = u \tag{4.11}$$

which is equivalent to

$$\frac{df}{dx} = 2xu. \tag{4.12}$$

When the internal parameter presents very short-range variations, some new structural patterns can be noticed. Considering an alternating input of the form $u = \sin(100x)$, it results in an observable physical quantity f represented in Figure 3; for an alternating cosine input represented by $u = \cos(100x)$, it results in the output f represented in Figure 4. Studying these

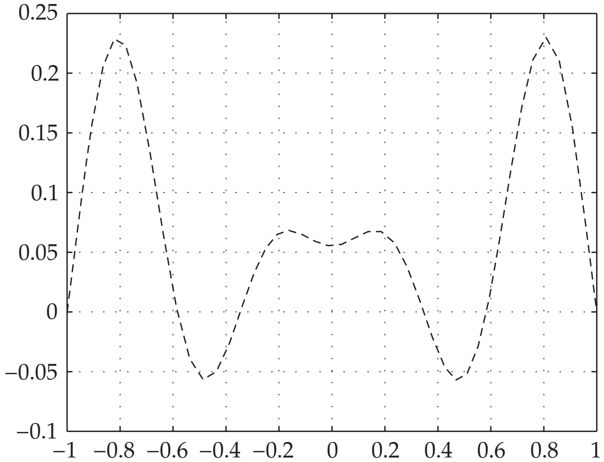


Figure 2: f versus distance for first-order system, input $u = \cos(10x)$.

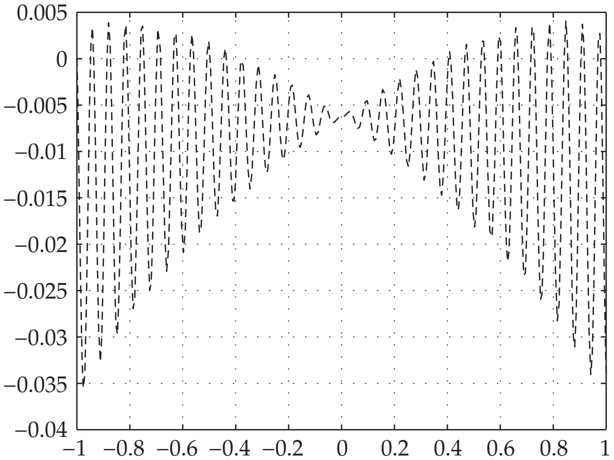


Figure 3: f versus distance for first-order system, input $u = \sin(100x)$.

two graphics, we can notice the presence of two distinct envelopes. Their shape depends on the phase of the input alternating component (the internal parameter), as related to the space origin. At first sight, an external observer could notice two distinct functions f inside the same material, along the Ox axis. These can be considered as two distinct structural patterns located in the same material, generated by a short-range alternating internal parameter u through a certain differential equation (invariant at the transformation $x \rightarrow -x$).

5. Aspects connected with short-range breaking phenomena

For simulating the generation of specific deformations inside a material medium under the action of external forces, it can be considered that some short wavelength vibrations appear in the area where the force acts. Usually the corresponding deformation is simulated inside the material medium, using linear differential equations or equations with partial derivatives (similar to the wave equation or to the equation of diffusion). Yet such linear equations cannot explain the distance between the space area where the external force acts and the space area

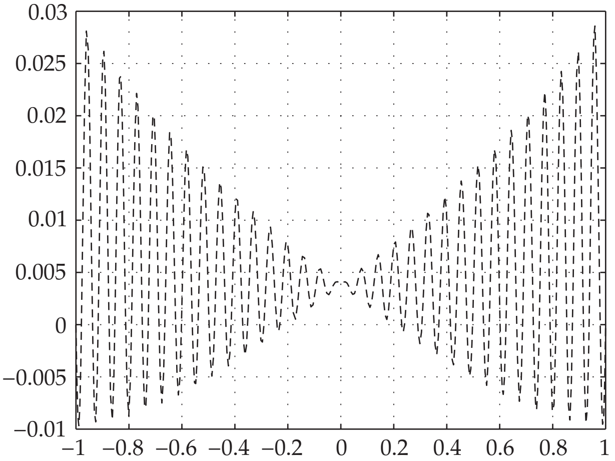


Figure 4: f versus distance for first-order system, input $u = \cos(100x)$.

where fracture phenomena appear. Using differential equations of higher order, some slow variations of deformation along a certain direction could be obtained. Due to the fact that the mathematical model should explain the sharp deformations at a certain distance of the point of space where the force acts (leading to fracture phenomena), some different types of differential equations must be studied. For this reason, our study has taken into consideration some dynamical equations able to generate practical test-functions (similar to wavelets) [1] and delayed pulses (when a free term which corresponds to an external pulse is added) [10] for justifying fracture phenomena appearing in a certain material medium. It is considered that an external force (described by a short wavelength sine function multiplied by a Gaussian function) acts upon the material medium in a certain area. As a consequence, some localized vibrations (corresponding to localized oscillations on closed space intervals presented in the previous paragraphs) appear. These localized oscillations are transmitted from one space interval to another according to a certain mathematical law which puts into correspondence the amplitude of these local vibrations to spatial coordinates.

Using a specific differential equation (able to generate symmetrical functions for a null free term) for describing the generation of the corresponding deformation along an axis inside the material medium, it results that a significant deformation appears at a certain distance. This significant deformation justifies the fracture phenomena, while the inner structure of the material cannot allow significant sharp deformations without breaking. The main problem is represented by the search of an adequate free term $u(x)$ able to justify fracture phenomena. We start by using a constant free term, using an equation as

$$f^{(2)} = \frac{0.6x^4 - 0.36x^2 - 0.2}{\left(x^2 - 1\right)^4} f + u(x), \tag{5.1}$$

where $u(x)$ represents the external force (supposed to be constant in a first approximation on the working space interval $(-1, 1)$). The deformation $f(x)$ is supposed to be first time generated by the external force at the limit $x = -1$ of the working interval and then (according to the differential equation) it generates the corresponding deformation along the whole working interval, with the external constant force u acting in a continuous manner upon the material. The deformation generated by such a constant force u should be symmetrical as related to

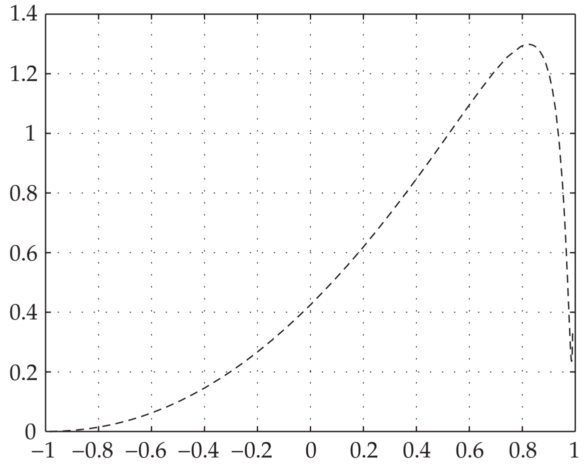


Figure 5: Deformation generated by an external constant force.

origin 0 (the previous differential equations being valid on the space interval $(-1,1)$ with initial null conditions for $f(x)$ at the initial point of space $x = -1$). The property of symmetry previously mentioned is justified by invariance properties of this type of differential equations [1]. However, even for $u(x) = 1$ (the most simple external force acting upon the material which is symmetrical as related to space origin 0) numerical simulations in MATLAB present an asymmetry of the output signal, justified by numerical errors (see Figure 5). But numerical simulations present also a slow varying deformation along the axis, with no spatial oscillations; thus the fracture phenomenon cannot be explained.

A similar shape of the output can be noticed for an input represented by a Gaussian external force, acting around the point of space $x = -0.9$ and having a width ten times smaller than the working period—similar to the use of a Gaussian modulated signal for generating delayed pulses [10]. In such a case the differential equation generating the deformation along the working interval is represented by

$$f^{(2)} = \frac{0.6x^4 - 0.36x^2 - 0.2}{(x^2-1)^4} f + \exp - \frac{(x+0.9)^2}{(0.01)^2} \tag{5.2}$$

and the corresponding output is represented in Figure 6.

So we must extend our search for adequate mathematical models, and we will try a free term $u(x)$ represented by

$$u(x) = \exp\left(-\frac{(x+0.9)^2}{(0.01)^2}\right) \sin 10^4 x. \tag{5.3}$$

This mathematical expression describes an external force represented by a Gaussian multiplied by a sine function with short wavelength, being considered that the applied force is transformed by the surface of the material into a set of alternating internal efforts with very short wavelength (similar to a localized vibration).

The corresponding output is represented in Figure 7. It can be noticed that we have finally obtained a sharp deformation appearing at a certain distance between the point of space

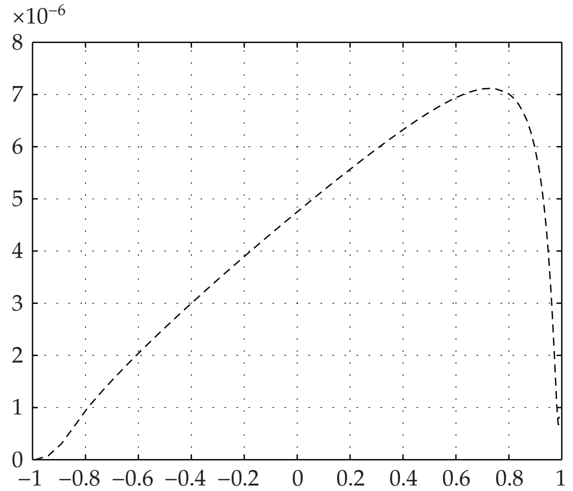


Figure 6: Deformation generated by an external Gaussian (localized) force.

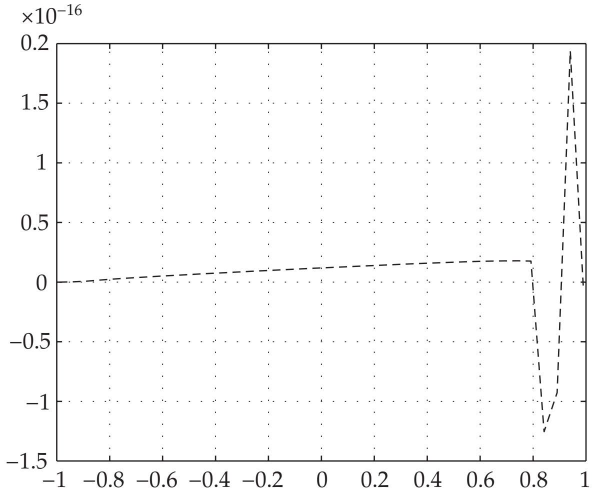


Figure 7: Deformation generated by a modulated Gaussian internal effort.

where the external (modulated Gaussian) force acts and the point of space where the sharp deformation appears. Moreover, the sharp deformation appears as an alternating function localized on a very short spatial interval. It is quite obvious that such a deformation cannot be allowed by the inner structure of the material, leading to fracture phenomena. This simulation explains also the fact that the fracture point is usually situated at a certain distance from the point where the external force is applied (as can be noticed studying the deformation presented in Figure 7 generated by the internal efforts $u(x)$ presented in Figure 8)

For the case when the Gaussian input is modulated by a cosine function, which means that

$$u(t) = \exp\left(- \frac{(\tau + 0.9)^2}{(0.01)^2} \right) \cos 10^4 \tau, \tag{5.4}$$

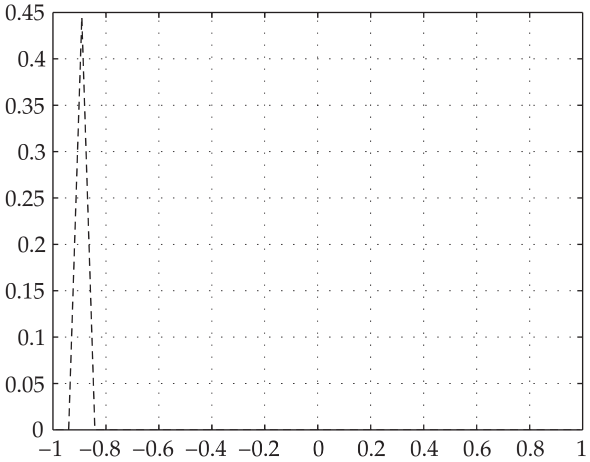


Figure 8: Modulated Gaussian internal effort by a sine function.

we obtain an output represented by a slowly varying function, without alternate deformation. So a cosine modulation of a Gaussian input is not suitable for simulating fracture phenomena appearing at a certain distance from the point where the external force acts.

We must point the fact that such localized alternating deformations generated by systems working on a limited interval and situated at a certain distance from the point where the external force acts differ to wavelets resulting from PDE equations (see [11]) and to propagating wavelets through dispersive media [12], while the shape of the resulting deformation is not symmetrical as related to *Ox* axis (its mean value differs to zero). However, a multiscale analysis of such pulses should be performed for explaining the complex fracture phenomena in an extended area and for justifying why a certain direction for generating deformation has to be chosen.

6. Conclusions

This study has shown that some solutions of the wave equation for half-closed space interval are considered around the point of space where the sources of the generated field are situated (e.g., the case of electrical charges generating the electromagnetic field). These solutions can be mathematically represented by vanishing waves corresponding to a superposition of traveling test-functions. Then some properties of spatial linear systems described by a certain physical quantity (generated by a differential equation) are studied. This quantity can be represented by internal electric or magnetic field inside the material or by similar physical quantities, and corresponds to the amplitude of localized oscillations previously mentioned. A specific mathematical law which can be approximated by a differential equation generates this quantity considering as input the spatial alternating variations of this internal parameter. As a consequence, specific spatial linear variations of the corresponding physical quantity appear. Finally, a specific differential equation (able to generate symmetrical functions for a null free term) is used for describing the generation of the corresponding deformation along an axis inside the material medium. Numerical simulations have shown that a significant deformation appears at a certain distance. This deformation justifies the fracture phenomena, while the inner structure of the material cannot allow significant sharp deformations without breaking.

References

[1] G. Toma, “Practical test-functions generated by computer algorithms,” in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA ’05)*, vol. 3482 of *Lecture Notes Computer Science*, pp. 576–584, Singapore, May 2005.

[2] J. F. M. Morgado and A. J. P. Gomes, “A derivative-free tracking algorithm for implicit curves with singularities,” in *Proceedings of the 4th International Conference in Computational Science (ICCS ’04)*, vol. 3039 of *Lecture Notes in Computer Science*, pp. 221–228, Krakow, Poland, June 2004.

[3] P. Federl and P. Prudinkiewiez, “Solving differential equations in developmental models of multicellular structures expressed using L-systems,” in *Proceedings of the 4th International Conference in Computational Science (ICCS ’04)*, vol. 3037 of *Lecture Notes in Computer Science*, pp. 65–72, Krakow, Poland, June 2004.

[4] C. Toma, “The possibility of appearing acausal pulses as solutions of the wave equation,” *The Hyperion Scientific Journal*, vol. 4, no. 1, pp. 25–29, 2004.

[5] L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Pergamon, New York, NY, USA, 3th edition, 1982.

[6] St. Pusca, “Invariance properties of practical test-functions used for generating asymmetrical pulses,” in *Proceedings of the International Conference Computational Science and Its Applications (ICCSA ’06)*, vol. 3980 of *Lecture Notes Computer Science*, pp. 763–770, Glasgow, UK, May 2006.

[7] Zh. Zhang, B. Ping, and W. Dong, “Oscillatory of unstable type second-order neutral difference equations,” *The Korean Journal of Computational & Applied Mathematics*, vol. 9, no. 1, pp. 87–99, 2002.

[8] J. Džurina, “Oscillation of second order differential equations with advanced argument,” *Mathematica Slovaca*, vol. 45, no. 3, pp. 263–268, 1995.

[9] F. Doboga, G. Toma, St. Pusca, M. Ghelmez, and C. Morarescu, “Filtering aspects of practical test-functions and the ergodic hypothesis,” in *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA ’05)*, vol. 3482 of *Lecture Notes Computer Science*, pp. 563–568, Singapore, May 2005.

[10] B. Lazar, A. Sterian, St. Pusca, V. Paun, C. Toma, and C. Morarescu, “Simulating delayed pulses in organic materials,” in *Proceedings of the International Conference Computational Science and Its Applications (ICCSA ’06)*, vol. 3980 of *Lecture Notes Computer Science*, pp. 779–784, Glasgow, UK, May 2006.

[11] C. Cattani, “Connection coefficients of Shannon wavelets,” *Mathematical Modelling and Analysis*, vol. 11, no. 2, pp. 117–132, 2006.

[12] C. Cattani and J. Rushchitsky, *Wavelet and Wave Analysis as Applied to Materials with Micro or Nanostructure*, vol. 74 of *Series on Advances in Mathematics for Applied Sciences*, World Scientific, Hackensack, NJ, USA, 2007.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 945420, 8 pages
doi:10.1155/2008/945420

Research Article

Solving Ratio-Dependent Predator-Prey System with Constant Effort Harvesting Using Homotopy Perturbation Method

Abdoul R. Ghotbi,[1] A. Barari,[2] and D. D. Ganji[2]

[1] *Department of Civil Engineering, Shahid Bahonar University of Kerman, Kerman 76169, Iran*
[2] *Department of Civil and Mechanical Engineering, Mazandaran University of Technology, P.O. Box 484, Babol 47144, Iran*

Correspondence should be addressed to Abdoul R. Ghotbi, civil_ghotbi40@yahoo.com

Received 1 February 2008; Revised 29 February 2008; Accepted 13 March 2008

Recommended by Cristian Toma

Due to wide range of interest in use of bioeconomic models to gain insight into the scientific management of renewable resources like fisheries and forestry, homotopy perturbation method is employed to approximate the solution of the ratio-dependent predator-prey system with constant effort prey harvesting. The results are compared with the results obtained by Adomian decomposition method. The results show that, in new model, there are less computations needed in comparison to Adomian decomposition method.

1. Introduction

Partial differential equations which arise in real-world physical problems are often too complicated to be solved exactly, and even if an exact solution is obtainable, the required calculations may be practically too complicated, or it might be difficult to interpret the outcome. Very recently, some promising approximate analytical solutions are proposed such as Exp-function method, Adomian decomposition method (ADM), variational iteration method (VIM), and homotopy perturbation method (HPM).

HPM is the most effective and convenient method for both linear and nonlinear equations. This method does not depend on a small parameter. Using homotopy technique in topology, a homotopy is constructed with an embedding parameter $p \in [0,1]$, which is considered as a "small parameter." HPM has been shown to effectively, easily, and accurately solve a large class of linear and nonlinear problems with components converging to accurate solutions. HPM was first proposed by He [1–7] and was successfully applied to various engineering problems.

The motivation of this paper is to extend the homotopy perturbation method (HPM) [8–17] to solve the ratio-dependent predator-prey system. The results of HPM are compared with those obtained by the ADM [18]. Different from ADM, where specific algorithms are usually used to determine the Adomian polynomials, HPM handles linear and nonlinear problems in simple manner by deforming a difficult problem into a simple one. The HPM is useful to obtain exact and approximate solutions of linear and nonlinear differential equations.

In this paper, we assume that the predator in model is not of commercial importance. The prey is subjected to constant effort harvesting with r, a parameter that measures the effort being spent by a harvesting agency. The harvesting activity does not affect the predator population directly. It is obvious that the harvesting activity does reduce the predator population indirectly by reducing the availability of the prey to the predator. Adopting a simple logistic growth for prey population with $e > 0$, $b > 0$, and $c > 0$ standing for the predator death rate, capturing rate, and conversion rate, respectively, we formulate the problem as

$$\begin{aligned} \frac{dx}{dt} &= x(1-x) - \frac{bxy}{y+x} - rx, \\ \frac{dy}{dt} &= \frac{cxy}{y+x} - ey, \end{aligned} \tag{1.1}$$

where $x(t)$ and $y(t)$ represent the fractions of population densities for prey and predator at time t, respectively. Equations (1.1) are to be solved according to biologically meaningful initial conditions $x(0) \geq 0$ and $y(0) \geq 0$ [18].

2. Applications

In this section, we will apply the HPM to nonlinear differential system of ratio-dependant predator-prey,

$$H(v,p) = (1-p)\left[L(v) - L(u_0)\right] + p\left[A(v) - f(r)\right] = 0, \quad p \in [0,1], \; r\varepsilon\Omega, \tag{2.1}$$

where $A(v)$ is a general differential operator which can be divided into a linear part $L(v)$ and a nonlinear part $N(v)$ and $f(r)$ is a known analytical function. $p \in [0,\ 1]$ is an embedding parameter, while u_0 is an initial approximation of the equation which should be solved, and satisfies the boundary conditions.

According to the HPM (relation (2.1)), we can construct a homotopy of system as follows:

$$\begin{aligned} &(1-p)\left(v_2\dot{v}_1 + v_1\dot{v}_1 - \dot{x}_0\mathrm{y}_0 - \dot{x}_0x_0\right) + p\left(v_2\dot{v}_1 + v_1\dot{v}_1 - (1-b-r)v_1v_2 + v_2v_1^2 - (1-r)v_1^2 + v_1^3\right) = 0, \\ &(1-p) \times \left(v_2\dot{v}_2 + v_1\dot{v}_2 - \dot{\mathrm{y}}_0\mathrm{y}_0 - \mathrm{x}_0\dot{\mathrm{y}}_0\right) + p\left(v_2\dot{v}_2 + v_1\dot{v}_2 + (e-c)v_1v_2 + \mathrm{e}v_2^2\right) = 0, \end{aligned} \tag{2.2}$$

where dot denotes differentiation with respect to t, and the initial approximations are as follows:

$$\begin{aligned} v_{1,0}(t) &= x_0(t) = x(0), \\ v_{2,0}(t) &= y_0(t) = y(0)\,. \end{aligned} \tag{2.3}$$

Assume that the solution of (2.2) can be written as a power series in p as follows:

$$\begin{aligned} \nu_1 &= \nu_{1,0} + p\nu_{1,1} + p^2\nu_{1,2} + p^3\nu_{1,3} + \cdots, \\ \nu_2 &= \nu_{2,0} + p\nu_{2,1} + p^2\nu_{2,2} + p^3\nu_{2,3} + \cdots, \end{aligned} \tag{2.4}$$

where $\nu_{i,j}$ $(i, j = 1, 2, 3, \ldots)$ are functions yet to be determined. Substituting (2.3) and (2.4) into (2.2), and arranging the coefficients of p powers, we have

$$\begin{aligned} &\left(v_{2,0}\dot{v}_{1,0} + v_{1,0}\dot{v}_{1,0}\right) \\ &\quad + \left(v_{1,0}^3 - v_{1,0}^2 + v_{1,0}\,\dot{v}_{1,1} + v_{2,0}\dot{v}_{1,1} + r\,v_{1,0}v_{2,0} + bv_{1,0}v_{2,0} - v_{1,0}v_{2,0} + v_{2,0}v_{1,0}^2 + rv_{1,0}^2\right)p \\ &\quad + \left(v_{1,1}\,\dot{v}_{1,1}+v_{1,0}\,\dot{v}_{1,2}+v_{2,0}\,\dot{v}_{1,2}+v_{2,1}\,\dot{v}_{1,1}+2r\,v_{1,0}v_{1,1} + b\,v_{1,0}v_{2,1} + 2\,v_{2,0}\,v_{1,0}\,v_{1,1} + r\,v_{1,1}\,v_{2,0}\right. \\ &\qquad\quad \left. + r\,v_{1,0}\,v_{2,1} + b\,v_{1,1}\,v_{2,0} - v_{1,0}\,v_{2,1}-v_{1,1}v_{2,0} + v_{2,1}v_{1,0}^2 - 2\,v_{1,0}\,v_{1,1} + 3\,v_{1,0}^2v_{1,1}\right)p^2 \\ &\quad + \left(v_{1,1}\dot{v}_{1,2} + v_{1,2}\dot{v}_{1,1} + v_{1,0}\dot{v}_{1,3} + v_{2,1}\dot{v}_{1,2} + v_{2,0}\dot{v}_{1,3} + v_{2,2}\dot{v}_{1,1} + v_{2,0}v_{1,1}^2 - v_{1,0}v_{2,2} - v_{1,2}v_{2,0}\right. \\ &\qquad\quad - v_{1,1}v_{2,1}+v_{2,2}v_{1,0}^2 + rv_{1,1}^2+3v_{1,0}v_{1,1}^2-v_{1,1}^2 + bv_{1,1}v_{2,1} + bv_{1,0}v_{2,2} + bv_{1,2}v_{2,0} + rv_{1,0}v_{2,2} \\ &\qquad\quad + rv_{1,1}v_{2,1} + rv_{1,2}v_{2,0} + 2v_{2,0}v_{1,0}v_{1,2} + 2rv_{1,0}v_{1,2} + 2v_{2,1}v_{1,0}v_{1,1} + 3v_{1,0}^2v_{1,2} \\ &\qquad\quad \left. - 2v_{1,0}v_{1,2}\right)p^3 + \cdots = 0, \\ &\left(v_{2,0}\dot{v}_{2,0} + v_{1,0}\dot{v}_{2,0}\right) \\ &\quad + \left(ev_{1,0}v_{2,0} - cv_{1,0}v_{2,0} + v_{2,0}\dot{v}_{2,1} + v_{1,0}\dot{v}_{2,1} + ev_{2,0}^2\right)p \\ &\quad + \left(v_{2,1}\dot{v}_{2,1}+ev_{1,0}v_{2,1}-cv_{1,0}v_{2,1}+ev_{1,1}v_{2,0}-cv_{1,1}v_{2,0}+2ev_{2,0}v_{2,1}+v_{2,0}\dot{v}_{2,2}+v_{1,1}\dot{v}_{2,1}+v_{1,0}\dot{v}_{2,2}\right)p^2 \\ &\quad + \left(ev_{2,1}^2 + v_{2,1}\dot{v}_{2,2} + v_{2,2}\dot{v}_{2,1} + v_{2,0}\dot{v}_{2,3} + v_{1,1}\dot{v}_{2,2} + v_{1,2}\dot{v}_{2,1} + v_{1,0}\dot{v}_{2,3} + ev_{1,0}v_{2,2} + ev_{1,1}v_{2,1}\right. \\ &\qquad\qquad \left. - cv_{1,0}v_{2,2} - cv_{1,1}v_{2,1} + ev_{1,2}v_{2,0} - cv_{1,2}v_{2,0} + 2ev_{2,0}v_{2,2}\right)p^3 + \cdots = 0. \end{aligned} \tag{2.5}$$

In order to obtain the unknown of $\nu_{i,j}(x, t)$, $i, j = 1, 2, 3, \ldots$, we must construct and solve the following system which includes 6 equations, considering the initial conditions of $\nu_{i,j}(0) = 0$, $i, j = 1, 2, 3, \ldots$:

$$\begin{aligned} &v_{2,0}\dot{v}_{1,0} + v_{1,0}\dot{v}_{1,0} = 0, \\ &v_{1,0}^3 - v_{1,0}^2 + v_{1,0}\,\dot{v}_{1,1} + v_{2,0}\dot{v}_{1,1} + v_{1,0}v_{2,0} + bv_{1,0}v_{2,0} - v_{1,0}v_{2,0} + v_{2,0}v_{1,0}^2 + rv_{1,0}^2 = 0, \\ &v_{1,1}\,\dot{v}_{1,1} + v_{1,0}\,\dot{v}_{1,2} + v_{2,0}\,\dot{v}_{1,2} + v_{2,1}\,\dot{v}_{1,1} + 2r\,v_{1,0}v_{1,1} + b\,v_{1,0}v_{2,1} + 2\,v_{2,0}\,v_{1,0}\,v_{1,1} + r\,v_{1,1}\,v_{2,0} \\ &\qquad + r\,v_{1,0}\,v_{2,1} + b\,v_{1,1}\,v_{2,0} - v_{1,0}\,v_{2,1} - v_{1,1}v_{2,0} + v_{2,1}v_{1,0}^2 - 2\,v_{1,0}\,v_{1,1} + 3\,v_{1,0}^2v_{1,1} = 0, \\ &v_{2,0}\dot{v}_{2,0} + v_{1,0}\dot{v}_{2,0} = 0, \\ &ev_{1,0}v_{2,0} - cv_{1,0}v_{2,0} + v_{2,0}\dot{v}_{2,1} + v_{1,0}\dot{v}_{2,1} + ev_{2,0}^2 = 0, \\ &v_{2,1}\dot{v}_{2,1} + ev_{1,0}v_{2,1} - cv_{1,0}v_{2,1} + ev_{1,1}v_{2,0} - cv_{1,1}v_{2,0} + 2ev_{2,0}v_{2,1} + v_{2,0}\dot{v}_{2,2} + v_{1,1}\dot{v}_{2,1} + v_{1,0}\dot{v}_{2,2} = 0. \end{aligned} \tag{2.6}$$

From (2.4), if the first three approximations are sufficient, then setting $p = 1$ yields the approximate solution of (1.1) to

$$\begin{aligned} x(t) &= \lim_{p\to 1} v_1(t) = \sum_{k=0}^{k=3} v_{1,k}(t), \\ y(t) &= \lim_{p\to 1} v_2(t) = \sum_{k=0}^{k=3} v_{2,k}(t). \end{aligned} \tag{2.7}$$

Therefore,

$$\mathrm{v}_{1,0}(t) = x_0(t) = x(0), \tag{2.8}$$

$$v_{1,1}(t) = -\frac{\mathrm{x}_0(x_0^2 - x_0 - y_0 + x_0y_0 + ry_0 + by_0 + rx_0)t}{x_0 + y_0}, \tag{2.9}$$

$$\begin{aligned} v_{1,2}(t) = \frac{1}{2(x_0+y_0)^3}((x_0t^2(&3y_0x_0^2 - x_0^2by_0 + 2x_0^3by_0 + 3x_0^4r + 6x_0^3y_0^2 - 3y_0^3x_0 + x_0^3r^2 - 9x_0^3y_0 \\ &+ 6x_0^4y_0 - 9x_0^2y_0^2 + 2y_0^3x_0^2 - 2x_0^3r - 2ry_0^3 - 2by_0^3 + b^2y_0^3 + r^2y_0^3 \\ &+ x_0^2by_0r + 3x_0ry_0^2b + y_0^2x_0eb + bx_0^2y_0e - bx_0^2y_0c - 3x_0by_0^2 \\ &+ 3x_0y_0^2 + 3y_0^3x_0r + 3y_0^3x_0b - 6x_0^2ry_0 + 2x_0^5 - 3x_0^4 + y_0^3 + 2ry_0^3b \\ &+ 9x_0^3ry_0 - 6x_0ry_0^2 + 9x_0^2y_0^2r + 5x_0^2y_0^2b + x_0^3 + 3x_0r^2y_0^2 + 3x_0^2r^2y_0))), \end{aligned} \tag{2.10}$$

$$v_{2,0}(t) = y_0(t) = y(0)\,, \tag{2.11}$$

$$v_{2,1}(t) = \frac{\mathrm{y}_0(-ex_0 + cx_0 - ey_0)t}{y_0 + x_0}, \tag{2.12}$$

$$\begin{aligned} v_{2,2}(t) = -\frac{1}{2(y_0+x_0)^3}((y_0t^2(&3y_0ex_0^2c + y_0^2cx_0e + 2ex_0^3c - cx_0^2y_0 - cx_0y_0^2 - c^2x_0^3 + cx_0^3y_0 \\ &+ cx_0^2y_0r + cx_0y_0^2b + cx_0^2y_0^2 + cx_0y_0^2r - e^2x_0^3 - 3y_0e^2x_0^2 \\ &- 3y_0^2e^2x_0 - y_0^3e^2))). \end{aligned} \tag{2.13}$$

We also obtained $v_{1,3}$ and $v_{2,3}$, but because they were too long to maintain, we skip them and only use them in the final numerical results. In this manner, the other components can be easily obtained by substituting (2.8) through (2.13) into (2.7) as follows:

$$\begin{aligned} x(t) = x(0) &- \left(\frac{\mathrm{x}_0(x_0^2 - x_0 - y_0 + x_0y_0 + ry_0 + by_0 + rx_0)t}{x_0 + y_0}\right) \\ &+ \frac{1}{2(x_0+y_0)^3}(x_0t^2(3y_0x_0^2 - x_0^2by_0 + 2x_0^3by_0 + 3x_0^4r + 6x_0^3y_0^2 - 3y_0^3x_0 + x_0^3r^2 - 9x_0^3y_0 \\ &\qquad + 6x_0^4y_0 - 9x_0^2y_0^2 + 2y_0^3x_0^2 - 2x_0^3r - 2ry_0^3 - 2by_0^3 + b^2y_0^3 + r^2y_0^3 \\ &\qquad + x_0^2by_0r + 3x_0ry_0^2b + y_0^2x_0eb + bx_0^2y_0e - bx_0^2y_0c - 3x_0by_0^2 + 3x_0y_0^2 \\ &\qquad + 3y_0^3x_0r + 3y_0^3x_0b - 6x_0^2ry_0 + 2x_0^5 - 3x_0^4 + y_0^3 + 2ry_0^3b + 9x_0^3ry_0 - 6x_0ry_0^2 \\ &\qquad + 9x_0^2y_0^2r + 5x_0^2y_0^2b + x_0^3 + 3x_0r^2y_0^2 + 3x_0^2r^2y_0)) + v_{1,3}\cdots, \\ y(t) = y(0) &+ \frac{\mathrm{y}_0(-ex_0 + cx_0 - ey_0)t}{y_0 + x_0} - \frac{1}{2(y_0+x_0)^3} \\ &\times (y_0t^2(3y_0ex_0^2c + y_0^2cx_0e + 2ex_0^3c - cx_0^2y_0 - cx_0y_0^2 - c^2x_0^3 + cx_0^3y_0 + cx_0^2y_0r \\ &\qquad + cx_0y_0^2b + cx_0^2y_0^2 + cx_0y_0^2r - e^2x_0^3 - 3y_0e^2x_0^2 - 3y_0^2e^2x_0 - y_0^3e^2)) + v_{2,3}\cdots. \end{aligned} \tag{2.14}$$

3. Numerical results and comparison with ADM

For comparison with the results obtained by ADM [18], the parameter values in four cases are considered in Table 1.

Table 1: Parameter values used for illustration purposes.

Case	x_0	y_0	b	c	e	r
1	0.5	0.3	0.8	0.2	0.5	0.9
2	0.5	0.3	0.8	0.2	0.5	0.1
3	0.5	0.6	0.5	0.5	0.3	0.1
4	0.5	0.2	0.5	0.5	0.1	0.2

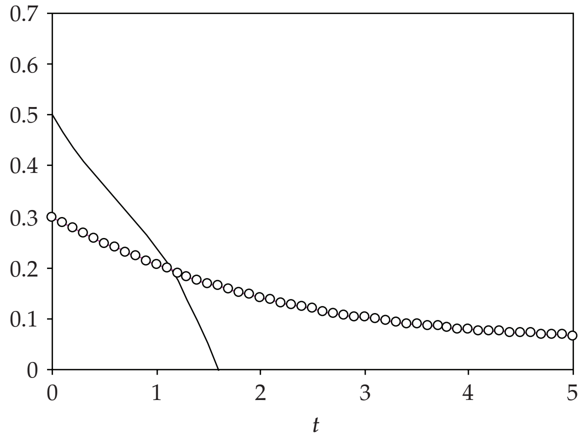


Figure 1: Population fraction versus time for Case 1: r = 0.9: (—) prey population fraction; (∘∘∘) predator population fraction.

Results of four terms approximation for $x(t), y(t)$ obtained by using HPM and ADM [18] are presented in (3.1), respectively:

$$
\begin{aligned}
\text{Case 1}: x &\approx 0.5 - 0.35t + 0.19476t^2 - 0.107288t^3,\\
y &\approx 0.3 - 0.1125t + 0.018808t^2 - 0.0011284t^3,\\
\text{Case 2}: x &\approx 0.5 + 0.05t + 0.012265t^2 - 0.0016032t^3,\\
y &\approx 0.3 - 0.1125t + 0.024433t^2 - 0.00398199t^3,\\
\text{Case 3}: x &\approx 0.3 + 0.0799\mathrm{t} + 0.00533\,\mathrm{t}^2 - 0.00115\,t^3,\\
y &\approx 0.6 - 0.08t + 0.01866t^2 - 0.00231t^3,\\
\text{Case 4}: x &\approx 0.5 + 0.07857t - 0.016020\,t^2 - 0.00119873\,t^3,\\
y &\approx 0.2 + 0.051428t + 0.0055918t^2 + 0.00002245t^3,\\
\text{Case 1}: x &\approx 0.5 - 0.35000t + 0.19476t^2 - 0.10728t^3,\\
y &\approx 0.3 - 0.11250t + 0.018809t^2 - 0.0011286t^3,\\
\text{Case 2}: x &\approx 0.5 + 0.05000t + 0.012266t^2 - 0.0016034t^3,\\
y &\approx 0.3 - 0.11250t + 0.024434t^2 - 0.0039821t^3,\\
\text{Case 3}: x &\approx 0.3 + 0.08000\mathrm{t} + 0.005333\,\mathrm{t}^2 - 0.0011555\,t^3,\\
y &\approx 0.6 - 0.08000t + 0.018667t^2 - 0.0023112t^3,\\
\text{Case 4}: x &\approx 0.5 + 0.07857t - 0.016021\,t^2 - 0.0011984t^3,\\
y &\approx 0.2 + 0.051430t + 0.0055920t^2 + 0.00002246t^3.
\end{aligned}
\tag{3.1}
$$

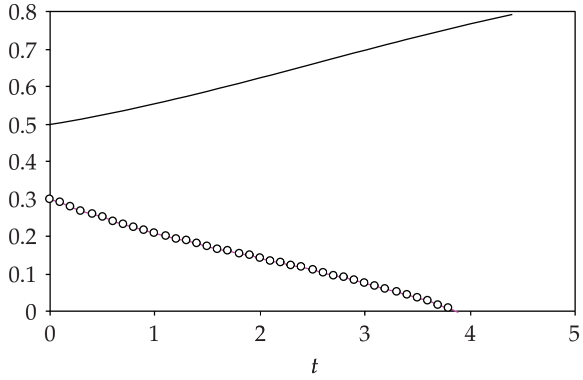


Figure 2: Population fraction versus time for Case 2: $r = 0.1$: (—) prey population fraction; (∘∘∘) predator population fraction.

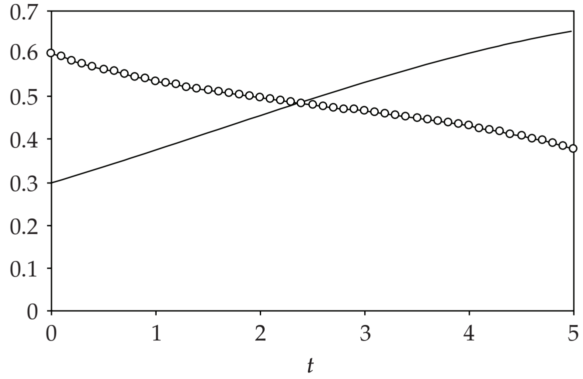


Figure 3: Population fraction versus time for Case 3: $r = 0.1$: (—) prey population fraction; (∘∘∘) predator population fraction.

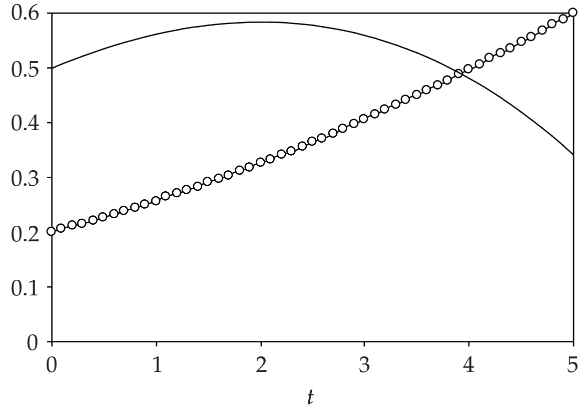


Figure 4: Population fraction versus time for Case 4: $r = 0.2$: (—) prey population fraction; (∘∘∘) predator population fraction.

Figures 1–4 show the relations between prey and predator populations versus time.

A noteworthy observation from Figure 1 is that prey and predator species can become extinct simultaneously for some values of parameters, regardless of the initial values. Thus, overexploitation of the prey population by constant effort harvesting process together with high predator capturing rate may lead to mutual extinction as a possible outcome of predator-pray interaction. In Figure 2, only the predator population gradually decreases and becomes extinct despite the availability of increasing prey population. This can be attributed to the effect of the predator death rate, being greater than the conversion rate and low constant prey harvesting as shown in Case 2 (see Table 1). Figures 3 and 4 illustrate the possibility of predator and prey long-term coexistence. Depending on the initial values, both prey and predator populations increase or reduce in order to allow long-term coexistence [18].

4. Conclusion

Homotopyperturbation method was employed to approximate the solution of the ratio-dependent predator-prey system with constant effort prey harvesting. The results obtained here were compared with results of Adomian decomposition method. The results show that there is less computations needed in comparison to ADM.

References

[1] J.-H. He, “New interpretation of homotopy perturbation method,” *International Journal of Modern Physics B*, vol. 20, no. 18, pp. 2561–2568, 2006.

[2] J.-H. He, “Some asymptotic methods for strongly nonlinear equations,” *International Journal of Modern Physics B*, vol. 20, no. 10, pp. 1141–1199, 2006.

[3] J.-H. He, “Homotopy perturbation method: a new nonlinear analytical technique,” *Applied Mathematics and Computation*, vol. 135, no. 1, pp. 73–79, 2003.

[4] J.-H. He, “A coupling method of a homotopy technique and a perturbation technique for non-linear problems,” *International Journal of Non-Linear Mechanics*, vol. 35, no. 1, pp. 37–43, 2000.

[5] J.-H. He, “A new approach to nonlinear partial differential equations,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 2, no. 4, pp. 230–235, 1997.

[6] J.-H. He, “Approximate solution of nonlinear differential equations with convolution product nonlinearities,” *Computer Methods in Applied Mechanics and Engineering*, vol. 167, no. 1-2, pp. 69–73, 1998.

[7] J.-H. He, “Homotopy perturbation technique,” *Computer Methods in Applied Mechanics and Engineering*, vol. 178, no. 3-4, pp. 257–262, 1999.

[8] M. Gorji, D. D. Ganji, and S. Soleimani, “New application of He’s homotopy perturbation method,” *International Journal of Nonlinear Sciences and Numerical Simulation*, vol. 8, no. 3, pp. 319–328, 2007.

[9] A. Sadighi and D. D. Ganji, “Solution of the generalized nonlinear Boussinesq equation using homotopy perturbation and variational iteration methods,” *International Journal of Nonlinear Sciences and Numerical Simulation*, vol. 8, no. 3, pp. 435–443, 2007.

[10] H. Tari, D. D. Ganji, and M. Rostamian, “Approximate solutions of K (2,2), KdV and modified KdV equations by variational iteration method, homotopy perturbation method and homotopy analysis method,” *International Journal of Nonlinear Sciences and Numerical Simulation*, vol. 8, no. 2, pp. 203–210, 2007.

[11] D. D. Ganji and A. Sadighi, “Application of He’s homotopy-perturbation method to nonlinear coupled systems of reaction-diffusion equations,” *International Journal of Nonlinear Sciences and Numerical Simulation*, vol. 7, no. 4, pp. 411–418, 2007.

[12] M. Rafei and D. D. Ganji, “Explicit solutions of Helmholtz equation and fifth-order KdV equation using homotopy perturbation method,” *International Journal of Nonlinear Sciences and Numerical Simulation*, vol. 7, no. 3, pp. 321–328, 2006.

[13] D. D. Ganji, “The application of He’s homotopy perturbation method to nonlinear equations arising in heat transfer,” *Physics Letters A*, vol. 355, no. 4-5, pp. 337–341, 2006.

[14] D. D. Ganji and A. Rajabi, "Assessment of homotopy-perturbation and perturbation methods in heat radiation equations," *International Communications in Heat and Mass Transfer*, vol. 33, no. 3, pp. 391–400, 2006.

[15] A. R. Ghotbi, M. A. Mohammadzade, A. Avaei, and M. Keyvanipoor, "A new approach to solve nonlinear partial differential equations," *Journal of Mathematics and Statistics*, vol. 3, no. 4, pp. 201–206, 2007.

[16] A. R. Ghotbi, A. Avaei, A. Barari, and M. A. Mohammadzade, "Assessment of He's homotopy perturbation method in Burgers and coupled Burgers' equations," *Journal of Applied Sciences*, vol. 8, no. 2, pp. 322–327, 2008.

[17] A. Barari, A. R. Ghotbi, F. Farrokhzad, and D. D. Ganji, "Variational iteration method and Homotopy-perturbation method for solving different types of wave equations," *Journal of Applied Sciences*, vol. 8, no. 1, pp. 120–126, 2008.

[18] O. D. Makinde, "Solving ratio-dependent predator-prey system with constant effort harvesting using Adomian decomposition method," *Applied Mathematics and Computation*, vol. 186, no. 1, pp. 17–22, 2007.