Advances in Intelligent Systems and Computing 240

Jerzy Świątek
Adam Grzech
Paweł Świątek
Jakub M. Tomczak *Editors*

Advances in Systems Science

Proceedings of the International Conference on Systems Science 2013 (ICSS 2013)

Springer

Advances in Intelligent Systems and Computing

Volume 240

Series Editor

Janusz Kacprzyk, Warsaw, Poland

For further volumes:
http://www.springer.com/series/11156

Jerzy Świątek · Adam Grzech
Paweł Świątek · Jakub M. Tomczak
Editors

Advances in Systems Science

Proceedings of the International Conference on Systems Science 2013 (ICSS 2013)

Springer

Editors
Jerzy Świątek
Institute of Informatics
Wroclaw University of Technology
Wrocław
Poland

Paweł Świątek
Institute of Informatics
Wroclaw University of Technology
Wrocław
Poland

Adam Grzech
Institute of Informatics
Wroclaw University of Technology
Wrocław
Poland

Jakub M. Tomczak
Institute of Informatics
Wroclaw University of Technology
Wrocław
Poland

ISSN 2194-5357 ISSN 2194-5365 (electronic)
ISBN 978-3-319-01856-0 ISBN 978-3-319-01857-7 (eBook)
DOI 10.1007/978-3-319-01857-7
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013946038

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The International Conference on Systems Science 2013 (ICSS 2013) was the 18th event of the series of international scientific conferences for researchers and practitioners in the fields of systems science and systems engineering. The conference took place in Wroclaw, Poland during September 10–12, 2013 and was organized by Wroclaw University of Technology and co-organized by: Committee of Automatics and Robotics of Polish Academy of Sciences, Committee of Computer Science of Polish Academy of Sciences and Polish Section of IEEE.

The first International Conference on Systems Science organized by Wroclaw University of Technology was held in 1974 and was organized every year till 1980 when Coventry Polytechnic, Coventry, UK started to co-organize parallel scientific events called International Conference on Systems Engineering (ICSE) every two years. In 1984 the Wright State University, Dayton, Ohio, USA joined to co-operate and organize International Conference on Systems Engineering. In 1990 the ICSE moved from Wright State University, Dayton, Ohio, USA to Nevada State University, Las Vegas, USA. Now, the International Conference on Systems Science is organized every three years in Wroclaw, by Wroclaw University of Technology and in the remaining years the International Conference on Systems Engineering is organized in Las Vegas by Nevada State University or in Coventry by Coventry University. The aim of the International Conference on Systems Science (ICSS) and the International Conference on Systems Engineering (ICSE) series was to provide an international forum for scientific research in systems science and systems engineering.

This year, we received almost 140 papers from 34 countries. Each paper was reviewed by at least two members of Program Committee or Board of Reviewers. Only 76 best papers were selected for oral presentation and publication in the International Conference on Systems Science 2013 proceedings. The final acceptance rate was 55%.

The papers included in the proceedings cover the following topics:

- Control Theory
- Databases and Data Mining
- Image and Signal Processing

- Machine Learning
- Modelling and Simulation
- Operational Research
- Service Science
- Time Series and System Identification

Accepted and presented papers highlight new trends and challenges in systems science and systems engineering. The presenters show how new research could lead to new and innovative applications. We do hope you will find these results useful and inspiring for your future research.

We would like to thank the Program Committee and Board of Reviewers, essential for reviewing the papers to ensure a high standard.

Jerzy Świątek

Organization

Program Committee

P. Albertos (Spain)
A.V. Balakrishnan (USA)
S. Bańka (Poland)
A. Bartoszewicz (Poland)
L. Borzemski (Poland)
V.N. Burkov (Russia)
K.J. Burnham (UK)
L.M. Camarinha-Matos (Portugal)
A. Grzech (Poland)
K. Hasegawa (Japan)
T. Hasegawa (Japan)
A. Isidori (Italy)
D.J.G. James (UK)
J. Józefczyk (Poland)
J. Kacprzyk (Poland)
T. Kaczorek (Poland)
R. Kaszyński (Poland)
L. Keviczky (Hungary)
G. Klir (USA)
P. Kontogiorgis (USA)
J. Korbicz (Poland)
G.L. Kovacs (Hungary)
V. Kucera (Czech Republic)
A.B. Kurzhanski (Russia)
H. Kwaśnicka (Poland)
N. Lavesson (Sweden)
B. Neumann (Germany)
J.J. Lee (Korea)
S.Y. Nof (USA)
W. Pedrycz (Canada)
F. Pichler (Austria)
G.P. Rao (India)
A. Rindos (USA)
L. Rutkowski (Poland)
E. Szczerbicki (Australia)
H. Selvaraj (USA)
R. Słowiński (Poland)
M. Sugisaka (Japan)
A. Sydow (Germany)
J. Świątek (Poland)
R. Tadeusiewicz (Poland)
Y. Takahara (Japan)
M. Thoma (Germany)
S.G. Tzafestas (Greece)
H. Unbehauen (Germany)
R. Vallee (France)
J. Węglarz (Poland)
L.A. Zadeh (USA)

Organizing Committee

Conference Chairman
Jerzy Świątek

Conference Co-Chairman
Adam Grzech

Financial Chairman
Paweł Świątek

Technical Chairmen
Krzysztof Brzostowski
Jarosław Drapała

Conference Secretary
Jakub M. Tomczak

Contents

Databases and Data Mining

Image and Signal Processing

Machine Learning

Modelling and Simulation

Operational Research

Service Science

Time Series and System Identification

Erratum

Decoupling Zeros of Positive Continuous-Time Linear Systems and Electrical Circuit

Tadeusz Kaczorek

Białystok University of Technology, Faculty of Electrical Engineering,
Wiejska 45D, 15-351 Bialystok Poland
kaczorek@isep.pw.edu.pl

Abstract. Necessary and sufficient conditions for the reachability and observability of the positive continuous-time linear systems are established. Definitions of the input-decoupling zeros, output-decoupling zeros and input-output decoupling zeros are proposed. Some properties of the decoupling zeros are discussed. Decoupling zeros of positive electrical circuits are also addressed.

Keywords: decoupling zeros, positive, continuous-time, linear system, positive electrical circuits, observability, reachability.

1 Introduction

In positive systems inputs, state variables and outputs take only non-negative values. Examples of positive systems are industrial processes involving chemical reactors, heat exchangers and distillation columns, storage systems, compartmental systems, water and atmospheric pollution models. A variety of models having positive linear behavior can be found in engineering, management science, economics, social sciences, biology and medicine, etc. An overview of state of the art in positive linear theory is given in the monographs [2, 3].

The notions of controllability and observability and the decomposition of linear systems have been introduced by Kalman [15, 16]. Those notions are the basic concepts of the modern control theory [1, 2, 4, 14, 17, 21]. They have been also extended to positive linear systems [2, 3].

The reachability and controllability to zero of standard and positive fractional discrete-time linear systems have been investigated in [9] and controllability and observability of electrical circuits in [6, 8, 10]. The decomposition of positive discrete-time linear systems has been addressed in [5]. The notion of decoupling zeros of standard linear systems have been introduced by Rosenbrock [17]. The zeros of linear standard discrete-time system have been addressed in [20] and zeros of positive continuous-time and discrete-time linear systems has been defined in [18, 19]. The decoupling zeros of positive discrete-time linear systems has been introduced in [7] and of positive continuous-time systems in [12, 13]. The positivity and reachability of fractional electrical circuits have been investigated in [8].

J. Świątek et al. (eds.), *Advances in Systems Science*,
Advances in Intelligent Systems and Computing 240,
DOI: 10.1007/978-3-319-01857-7_1, © Springer International Publishing Switzerland 2014

In this paper the notions of decoupling zeros will be extended for positive continuous-time linear systems and electrical circuits.

The paper is organized as follows. In section 2 the basic definitions and theorems concerning reachability and observability of positive continuous-time linear systems are given. The decomposition of the pair (A,B) and (A,C) of positive linear system is addressed in section 3. The main result of the paper is given in section 4 where the definitions of the decoupling-zeros are proposed. The positive electrical circuits are addressed in section 5 and decoupling zeros of positive electrical circuits in section 6. Concluding remarks are given in section 7.

The following notation will be used: $\Re$ - the set of real numbers, $\Re^{n\times m}$ - the set of $n\times m$ real matrices, $\Re_+^{n\times m}$ - the set of $n\times m$ matrices with nonnegative entries and $\Re_+^n = \Re_+^{n\times 1}$, M_n - the set of $n\times n$ Metzler matrices (real matrices with nonnegative off-diagonal entries), I_n - the $n\times n$ identity matrix.

2 Reachability and Observability of Positive Continuous-Time Linear Systems

2.1 Reachability of Positive Systems

Consider the linear continuous-time system

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{2.1}$$

where $x(t)\in\Re^n$, $u(t)\in\Re^m$, $y(t)\in\Re^p$ are the state, input and output vectors and $A\in\Re^{n\times n}$, $B\in\Re^{n\times m}$, $C\in\Re^{p\times n}$, $D\in\Re^{p\times m}$.

Definition 2.1. [2, 3] The linear system (2.1) is called (internally) positive if $x(t)\in\Re_+^n$ and $y(t)\in\Re_+^p$, $t\ge 0$ for any $x(0)=x_0\in\Re_+^n$ and every $u(t)\in\Re_+^m$, $t\ge 0$.

Theorem 2.1. [2, 3] The system (2.1) is positive if and only if

$$A\in M_n,\quad B\in\Re_+^{n\times m},\quad C\in\Re_+^{p\times n},\quad D\in\Re_+^{p\times m} \tag{2.2}$$

Definition 2.2. The positive system (2.1) (or positive pair (A,B)) is called reachable at time t_f if for any given final state $x_f\in\Re_+^n$ there exists an input sequence $u(t)\in\Re_+^m$, $t\in[0,t_f]$ which steers the state of the system from zero state ($x(0)=0$) to state $x_f\in\Re_+^n$, i.e. $x(t_f)=x_f$.

A column $a \in \Re_+^n$ (row $a^T \in \Re_+^n$) is called monomial if only one its entry is positive and the remaining entries are zero. A real matrix $A \in \Re_+^{n \times n}$ is called monomial if each its row and each its column contains only one positive entry and the remaining entries are zero.

Theorem 2.2. The positive system (2.1) is reachable at time $t \in [0, t_f]$ if and only if the matrix $A \in M_n$ is diagonal and the matrix $B \in \Re_+^{n \times n}$ is monomial.
Proof is given in [12].

2.2 Observability of Positive Systems

Consider the positive system

$$\dot{x}(t) = Ax(t) \tag{2.3a}$$

$$y(t) = Cx(t) \tag{2.3b}$$

where $x(t) \in \Re_+^n$, $y(t) \in \Re_+^p$ and $A \in M_n$, $C \in \Re_+^{p \times n}$.

Definition 2.3. The positive system (2.3) is called observable if knowing the output $y(t) \in \Re_+^p$ and its derivatives $y^{(k)}(t) = \dfrac{d^k y(t)}{dt^k} \in \Re_+^p$, $k = 1,2,\ldots,n-1$ it is possible to find the initial values $x_0 = x(0) \in \Re_+^n$ of $x(t) \in \Re_+^n$.

Theorem 2.3. The positive system (2.3) is observable if and only if the matrix $A \in M_n$ is diagonal and the matrix

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \tag{2.4}$$

has n linearly independent monomial rows.
Proof is given in [12].

3 Decomposition of the Pairs (A,B) and (A,C)

3.1 Decomposition of the Pair (A,B)

Consider the pair (A,B) with A being diagonal

$$A = \operatorname{diag}[a_{11}, a_{22}, \ldots, a_{n,n}] \in M_n \tag{3.1a}$$

and the matrix B with m linearly independent columns $B_1, B_2, ..., B_m$

$$B = [B_1 \quad B_2 \quad ... \quad B_m]. \tag{3.1b}$$

By Theorem 2.2 the pair (3.1) is unreachable if $m < n$.

It will be shown that in this case the pair can be decomposed into the reachable pair $(\overline{A}_1, \overline{B}_1)$ and unreachable pair $(\overline{A}_2, \overline{B}_2 = 0)$.

Theorem 3.1. For the unreachable pair (3.1) ($m < n$) there exists a monomial matrix $P \in \Re_+^{n \times n}$ such that the pair (A,B) can be reduced to the form

$$\overline{A} = PAP^{-1} = \begin{bmatrix} \overline{A}_1 & 0 \\ 0 & \overline{A}_2 \end{bmatrix}, \quad \overline{B} = PB = \begin{bmatrix} \overline{B}_1 \\ 0 \end{bmatrix} \tag{3.2}$$

where $\overline{A}_1 = \text{diag}[\overline{a}_{11}, \overline{a}_{22}, ..., \overline{a}_{n_1,n_1}] \in M_{n_1}$, $\overline{A}_2 = \text{diag}[\overline{a}_{n_1+1,n_1+1}, ..., \overline{a}_{n,n}] \in M_{n_2}$, $\overline{B}_1 \in \Re_+^{n_1 \times m}$, $n = n_1 + n_2$, the pair $(\overline{A}_1, \overline{B}_1)$ is reachable and the pair $(\overline{A}_2, \overline{B}_2 = 0)$ is unreachable.

Proof. Performing on the matrix B the following elementary row operations:

- interchange the i-th and j-th rows, denoted by $L[i, j]$,
- multiplication of i-th rows by positive number c, denoted by $L[i \times c]$,

we may reduce the matrix B to the form $\begin{bmatrix} \overline{B}_1 \\ 0 \end{bmatrix}$, where $\overline{B}_1 \in \Re_+^{n_1 \times m}$ is monomial with positive entries equal to 1. Performing the same elementary row operations on the identity matrix In we obtain the desired monomial matrix P. It is well-known [3] that $P^{-1} \in \Re_+^{n \times n}$ and for diagonal matrix A we have $\overline{A} = PAP^{-1} = \begin{bmatrix} \overline{A}_1 & 0 \\ 0 & \overline{A}_2 \end{bmatrix}$. □

Example 3.1. Consider the unreachable pair (3.1) with

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 3 \\ 0 & 0 \\ 2 & 0 \end{bmatrix}. \tag{3.3}$$

Performing on the matrix B the following elementary row operations $L[1,3]$, $L[1 \times 1/2]$, $L[2,3]$, $L[2 \times 1/3]$ we obtain

$$\overline{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{3.4}$$

Performing the same elementary row operations on the identity matrix I_3 we obtain the desired monomial matrix

$$P = \begin{bmatrix} 0 & 0 & 1/2 \\ 1/3 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{3.5}$$

and

$$\begin{aligned} PB &= \begin{bmatrix} 0 & 0 & 1/2 \\ 1/3 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 0 & 0 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \overline{B} = \begin{bmatrix} \overline{B}_1 \\ 0 \end{bmatrix}, \\ \overline{A} &= PAP^{-1} = \begin{bmatrix} 0 & 0 & 1/2 \\ 1/3 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 3 & 0 \\ 0 & 0 & 1 \\ 2 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix} = \begin{bmatrix} \overline{A}_1 & 0 \\ 0 & \overline{A}_2 \end{bmatrix}. \end{aligned} \tag{3.6}$$

The positive pair

$$\overline{A}_1 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \overline{B}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.7}$$

is reachable and the pair $(\overline{A}_2, 0)$ is unreachable.

3.2 Decomposition of the Pair (*A*,*C*)

Let the observability matrix

$$O_n = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \in \Re_+^{pn \times n} \tag{3.8}$$

of the positive unobservable system has $n_1 < n$ linearly independent monomial rows.

If the conditions

$$Q_k A Q_j^T = 0 \text{ for } k = 1,2,...,n_1 \text{ and } j = n_1 + 1,...,n \tag{3.9}$$

are satisfied then there exists the monomial matrix [5, 6]

$$Q^T = [Q_{j_1}{}^T \quad ... \quad Q_{j_1\bar{d}_1}{}^T \quad Q_{j_2}^T \quad ... \quad Q_{j_2\bar{d}_2}^T \quad ... \quad Q_{j_l\bar{d}_l}^T \quad Q_{n_1+1}^T \quad ... \quad Q_n^T]^T \in \Re_+^{n \times n} \tag{3.10a}$$

where

$$Q_{j_1} = C_{j_1},...,Q_{j_1\bar{d}_1} = C_{j_1} A^{\bar{d}_1 - 1}, Q_{j_2} = C_{j_2},...,Q_{j_2\bar{d}_2} = C_{j_2} A^{\bar{d}_2 - 1},...,Q_{j_l\bar{d}_l} = C_{j_l} A^{\bar{d}_l - 1} \tag{3.10b}$$

and $\bar{d}_j, \quad j = 1,...,l$ are some natural numbers.

Theorem 3.2. Let the positive system (2.3) be unobservable and let there exist the monomial matrix (3.10). Then the pair (A,C) of the system can be reduced by the use of the matrix (3.10) to the form

$$\hat{A} = QAQ^{-1} = \begin{bmatrix} \hat{A}_1 & 0 \\ 0 & \hat{A}_2 \end{bmatrix}, \quad \hat{C} = CQ^{-1} = [\hat{C}_1 \quad 0]$$
$$\hat{A}_1 \in \Re_+^{n_1 \times n_1}, \quad \hat{A}_2 \in \Re_+^{n_2 \times n_2}, \quad (n_2 = n - n_1), \quad \hat{C}_1 \in \Re_+^{p \times n_1} \tag{3.11}$$

where the pair $(\hat{A}_1, \hat{C}_1)$ is observable and the pair $(\hat{A}_2, \hat{C}_2 = 0)$ is unobservable.

Proof is given in [5].

Example 3.2. Consider the unobservable pair

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad C = [0 \quad 0 \quad 1]. \tag{3.12}$$

In this case the observability matrix

$$Q_3 = \begin{bmatrix} C \\ CA \\ CA^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.13}$$

has only one monomial row $Q_1 = C$, i.e. $n_1 = 1$ and the conditions (3.9) are satisfied for $Q_2 = [1 \quad 0 \quad 0]$ and $Q_3 = [0 \quad 1 \quad 0]$ since $Q_1 A Q_j^T = 0$ for $j = 2,3$. The matrix (3.10) has the form

$$Q = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{3.14}$$

Using (3.11) and (3.14) we obtain

$$\hat{A} = QAQ^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix} = \begin{bmatrix} \hat{A}_1 & 0 \\ 0 & \hat{A}_2 \end{bmatrix},$$
$$\hat{C} = CQ^{-1} = [0 \quad 0 \quad 1] \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = [1 \quad 0 \quad 0] = [\hat{C}_1 \quad 0], \tag{3.15}$$

where

$$\hat{A}_1 = [-1], \quad \hat{A}_2 = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}, \quad \hat{C}_1 = [1]. \tag{3.16}$$

The pair $(\hat{A}_1, \hat{C}_1)$ is observable and the pair $(\hat{A}_2, 0)$ is unobservable.

4 Decoupling Zeros of the Positive Systems

It is well-known [17] that for standard linear systems the input-decoupling zeros are the eigenvalues of the matrix $\overline{A}_2$ of the unreachable (uncontrollable) part $(\overline{A}_2, \overline{B}_2 = 0)$.

In a similar way we will define the input-decoupling zeros of the positive continuous-time linear systems.

Definition 4.1. Let $\overline{A}_2$ be the matrix of unreachable part of the system (2.1). The zeros $s_{i1}, s_{i2}, ..., s_{i\overline{n}_2}$ of the characteristic polynomial

$$\det[I_{\overline{n}_2} s - \overline{A}_2] = s^{\overline{n}_2} + \overline{a}_{\overline{n}_2 - 1} s^{\overline{n}_2 - 1} + ... + \overline{a}_1 s + \overline{a}_0 \tag{4.1}$$

of the matrix $\overline{A}_2$ are called the input-decoupling zeros of the positive system (2.1). The list of the input-decoupling zeros will be denoted by $Z_i = \{s_{i1}, s_{i2}, ..., s_{i\overline{n}_2}\}$.

Theorem 4.1. The state vector $x(t)$ of the positive system (2.1) is independent of the input-decoupling zeros for any input $u(t)$ and zero initial conditions.

Proof. From (2.1) for zero initial conditions $x(0) = 0$ we have

$$X(s) = [I_n s - A]^{-1} B U(s), \tag{4.2}$$

where $X(s)$ and $U(s)$ are Laplace transforms of $x(t)$ and $u(t)$, respectively. Taking into account (3.2) we obtain

$$\begin{aligned} X(s) &= [I_n s - P^{-1}\overline{A}P]^{-1} P^{-1} B U(s) = P^{-1}[I_n s - \overline{A}]^{-1} \overline{B} U(s) \\ &= P^{-1} \begin{bmatrix} I_{\overline{n}_1} s - \overline{A}_1 & 0 \\ 0 & I_{\overline{n}_2} s - \overline{A}_2 \end{bmatrix}^{-1} \begin{bmatrix} \overline{B}_1 \\ 0 \end{bmatrix} U(s) = P^{-1} \begin{bmatrix} [I_{\overline{n}_1} s - \overline{A}_1]^{-1} \overline{B}_1 \\ 0 \end{bmatrix} U(s). \end{aligned} \tag{4.3}$$

From (4.3) it follows that $X(s)$ is independent of the matrix $\overline{A}_2$ and of the input-decoupling zeros for any input $u(t)$. □

Example 4.1. (continuation of Example 3.1) In Example 3.1 it was shown that for the unreachable pair (3.1) the matrix $\overline{A}_2$ has the form $\overline{A}_2 = [-2]$. Therefore, the positive system (3.1) with (3.3) has one input-decoupling zero $s_{i1} = -2$.

For standard continuous-time linear systems the output-decoupling zeros are defined as the eigenvalues of the matrix of the unobservable part of the system. In a similar way we will define the output-decoupling zeros of the positive continuous-time linear systems.

Definition 4.2. Let $\hat{A}_2$ be the matrix of unobservable part of the system (2.3). The zeros $s_{o1}, s_{o2},..., s_{o\hat{n}_2}$ of the characteristic polynomial

$$\det[I_{\hat{n}_2} s - \hat{A}_2] = s^{\hat{n}_2} + \hat{a}_{\hat{n}_2 - 1} s^{\hat{n}_2 - 1} + ... + \hat{a}_1 s + \hat{a}_0 \tag{4.4}$$

of the matrix $\hat{A}_2$ are called the output-decoupling zero of the positive system (2.3).

The list of the output-decoupling zeros will be denoted by $Z_o = \{s_{o1}, s_{o2},..., s_{o\hat{n}_2}\}$.

Theorem 4.2. The output vector $y(t)$ of the positive system (2.3) is independent of the output-decoupling zeros for any input $\overline{u}(t) = Bu(t)$ and zero initial conditions.

Proof is similar to the proof of Theorem 4.1.

Example 4.2. (continuation of Example 3.2) In Example 3.2 it was shown that the matrix $\hat{A}_2$ of the positive unobservable pair (3.12) has the form

$$\hat{A}_2 = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \tag{4.5}$$

and the positive system has two output-decoupling zero $s_{o1} = -1$, $s_{o2} = -2$.

Following the same way as for standard continuous-time linear systems we define the input-output decoupling zeros of the positive systems as follows.

Definition 4.3. Zeros $s_{io}^{(1)}, s_{io}^{(2)},..., s_{io}^{(k)}$ which are simultaneously the input-decoupling zeros and the output-decoupling zeros of the positive system are called the input-output decoupling zeros of the positive system, i.e.

$$s_{io}^{(j)} \in Z_i \text{ and } s_{io}^{(j)} \in Z_o \text{ for } j = 1,2,\ldots,k;\ k \le \min(\overline{n}_2, \hat{n}_2). \tag{4.6}$$

The list of input-output decoupling zeros will be denoted by $Z_{io} = \{z_{io}^{(1)}, z_{io}^{(2)},..., z_{io}^{(k)}\}$.

Example 4.3. Consider the positive system with the matrices *A*, *B*, *C* given by (3.3) and (3.12). In Example 4.1 it was shown that the positive system has one input-decoupling zero $s_{i1} = -2$ and in Example 4.2 that the system has two output-decoupling zeros $s_{o1} = -1, s_{o2} = -2$. Therefore, by Definition 4.3 the positive system has one input-output decoupling zero $s_{io}^{(1)} = -2$.

5 Positive Electrical Circuits

Example 5.1. Consider the electrical circuit shown in Figure 1 with given conductances $G_1, G'_1, G_2, G'_2, G_{12}$, capacitances C_1, C_2 and source voltages e_1, e_2.

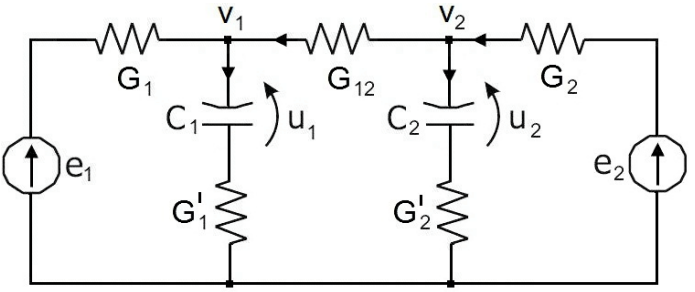

Fig. 1. Electrical circuit

Using the Kirchhoff's laws we can write the equations

$$\frac{d}{dt}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \frac{G'_1}{C_1} & 0 \\ 0 & \frac{G'_2}{C_2} \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} - \begin{bmatrix} \frac{G'_1}{C_1} & 0 \\ 0 & \frac{G'_2}{C_2} \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{5.1}$$

and

$$\begin{bmatrix} -G_{11} & G_{12} \\ G_{12} & -G_{22} \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = -\begin{bmatrix} G'_1 & 0 \\ 0 & G'_2 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} - \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \tag{5.2a}$$

where

$$G_{11} = G_1 + G'_1 + G_{12}, \quad G_{22} = G_2 + G'_2 + G_{12}. \tag{5.2b}$$

Taking into account that the matrix

$$\begin{bmatrix} -G_{11} & G_{12} \\ G_{12} & -G_{22} \end{bmatrix} \tag{5.3}$$

is nonsingular and

$$-\begin{bmatrix} -G_{11} & G_{12} \\ G_{12} & -G_{22} \end{bmatrix}^{-1} \in \Re_+^{2\times 2} \tag{5.4}$$

from (5.2a) we obtain

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = -\begin{bmatrix} -G_{11} & G_{12} \\ G_{12} & -G_{22} \end{bmatrix}^{-1}\left\{\begin{bmatrix} G'_1 & 0 \\ 0 & G'_2 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}\right\} \tag{5.5}$$

Substitution of (5.5) into (5.1) yields

$$\frac{d}{dt}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = A\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + B\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \tag{5.6}$$

where

$$A = -\begin{bmatrix} \frac{G'_1}{C_1} & 0 \\ 0 & \frac{G'_2}{C_2} \end{bmatrix}\begin{bmatrix} -G_{11} & G_{12} \\ G_{12} & -G_{22} \end{bmatrix}^{-1}\begin{bmatrix} G'_1 & 0 \\ 0 & G'_2 \end{bmatrix} - \begin{bmatrix} \frac{G'_1}{C_1} & 0 \\ 0 & \frac{G'_2}{C_2} \end{bmatrix} \in M_2, \tag{5.7a}$$

$$B = -\begin{bmatrix} \frac{G'_1}{C_1} & 0 \\ 0 & \frac{G'_2}{C_2} \end{bmatrix}\begin{bmatrix} -G_{11} & G_{12} \\ G_{12} & -G_{22} \end{bmatrix}^{-1}\begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \in \Re_+^{2\times 2}. \tag{5.7b}$$

From (5.7) it follows that A is Metzler matrix and the matrix B has nonnegative entries. Therefore, the electrical circuit is positive for all values of the conductances and capacitances.

In general case we have the following theorem.

Theorem 5.1. The electrical circuit shown in Figure 2 is positive for all values of the conductances, capacitances and source voltages.
Proof is given in [8, 10].

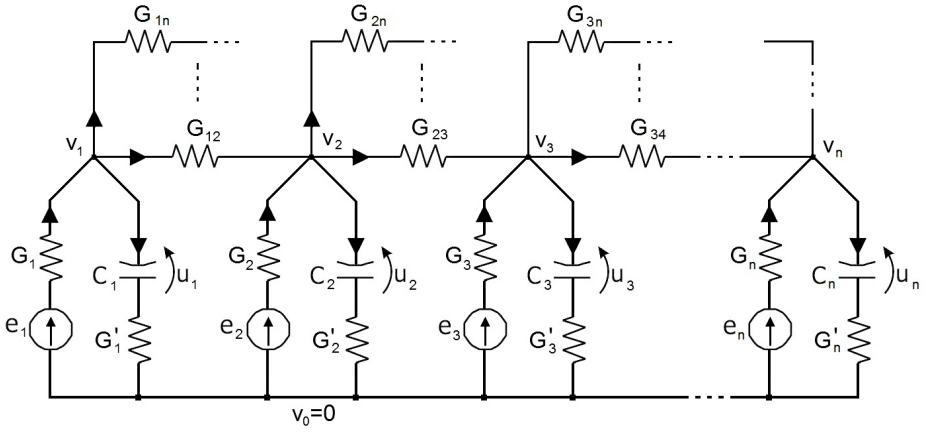

Fig. 2. Electrical circuit

Note that the standard electrical circuit shown in Figure 5.2 is reachable for all nonzero values of the conductances and capacitances since $\det B \neq 0$.

Theorem 5.2. The electrical circuit shown in Figure 5.2 is reachable if and only if

$$G_{k,j} = 0 \quad \text{for} \quad k \neq j \quad \text{and} \quad k, j = 1,...n.. \tag{5.8}$$

Proof. It is easy to see that the matrices $A \in M_n$ and $B \in \Re_+^{n \times n}$ are both diagonal matrices if and only if the condition (5.8) is satisfied. In this case by Theorem 2.2 the electrical circuit is reachable if and only if the conditions (5.8) are met. □

Example 5.2. Consider the electrical circuit shown in Figure 3 with given resistances R_1, R_2, R_3, inductances L_1, L_2, L_3 and source voltages e_1, e_3.

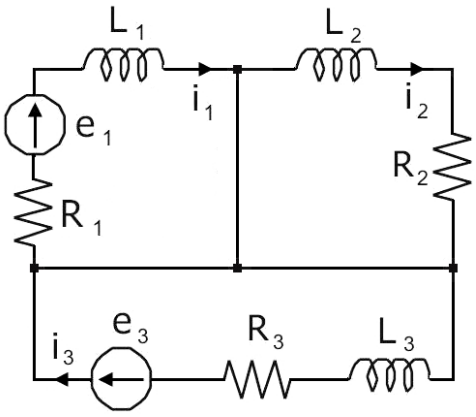

Fig. 3. Electrical circuit

Using the Kirchhoff's laws we can write the equations

$$\begin{aligned} L_1 \frac{di_1}{dt} &= -R_1 i_1 + e_1 \\ L_2 \frac{di_2}{dt} &= -R_2 i_2 \\ L_3 \frac{di_3}{dt} &= -R_3 i_3 + e_3 \end{aligned} \tag{5.9}$$

which can be written in the form

$$\frac{d}{dt}\begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix} = A \begin{bmatrix} i_1 \\ i_2 \\ i_3 \end{bmatrix} + B \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \tag{5.10a}$$

where

$$A = \begin{bmatrix} -\frac{R_1}{L_1} & 0 & 0 \\ 0 & -\frac{R_2}{L_2} & 0 \\ 0 & 0 & -\frac{R_3}{L_3} \end{bmatrix}, \quad B = \begin{bmatrix} \frac{1}{L_1} & 0 \\ 0 & 0 \\ 0 & \frac{1}{L_3} \end{bmatrix}. \tag{5.10b}$$

By Theorem 2.2 the positive electrical circuit (or the pair (5.10b)) is unreachable since $n = 3 < m = 2$.

The unreachable pair (5.10b) can be decomposed into reachable pair $(\overline{A}_1, \overline{B}_1)$ and unreachable pair $(\overline{A}_2, \overline{B}_2 = 0)$

In this case the monomial matrix P has the form

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{5.11}$$

and we obtain

$$\overline{B} = PB = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{L_1} & 0 \\ 0 & 0 \\ 0 & \frac{1}{L_3} \end{bmatrix} = \begin{bmatrix} \frac{1}{L_1} & 0 \\ 0 & \frac{1}{L_3} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \overline{B}_1 \\ 0 \end{bmatrix}, \quad \overline{B}_1 = \begin{bmatrix} \frac{1}{L_1} & 0 \\ 0 & \frac{1}{L_3} \end{bmatrix},$$

$$\overline{A} = PAP^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -\frac{R_1}{L_1} & 0 & 0 \\ 0 & -\frac{R_2}{L_2} & 0 \\ 0 & 0 & -\frac{R_3}{L_3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} -\frac{R_1}{L_1} & 0 & 0 \\ 0 & -\frac{R_3}{L_3} & 0 \\ 0 & 0 & -\frac{R_2}{L_2} \end{bmatrix} = \begin{bmatrix} \overline{A}_1 & 0 \\ 0 & \overline{A}_2 \end{bmatrix}. \tag{5.12}$$

and

$$\overline{A}_1 = \begin{bmatrix} -\frac{R_1}{L_1} & 0 \\ 0 & -\frac{R_3}{L_3} \end{bmatrix}, \quad \overline{A}_2 = \begin{bmatrix} -\frac{R_2}{L_2} \end{bmatrix}. \tag{5.13}$$

The reachable pair $(\overline{A}_1, \overline{B}_1)$ is reachable and the pair $(\overline{A}_2, \overline{B}_2 = 0)$ is unreachable.

In general case we have the following theorem [10].

Theorem 5.3. The linear electrical circuit composed of resistors, coils and voltage source is positive for any values of the resistances, inductances and source voltages if the number of coils is less or equal to the number of its linearly independent meshes and the direction of the mesh currents are consistent with the directions of the mesh source voltages.

These considerations can be extended to fractional positive electrical circuits [8].

6 Decoupling Zeros of the Positive Electrical Circuits

In a similar way as for linear systems we will define the input-decoupling zeros of the positive electrical circuits.

Definition 6.1. Let $\overline{A}_2$ be the matrix of unreachable part of the positive electrical circuit (2.1). The zeros $s_{i1}, s_{i2},..., s_{i\overline{n}_2}$ of the characteristic polynomial

$$\det[I_{\overline{n}_2} s - \overline{A}_2] = s^{\overline{n}_2} + \overline{a}_{\overline{n}_2 - 1} s^{\overline{n}_2 - 1} + ... + \overline{a}_1 s + \overline{a}_0 \tag{6.1}$$

of the matrix $\overline{A}_2$ are called the input-decoupling zeros of the positive electrical circuit (2.1).

The list of the input-decoupling zeros will be denoted by $Z_i = \{s_{i1}, s_{i2},..., s_{in_2}\}$.

Theorem 6.1. The state vector $x(t)$ of the positive electrical circuit is independent of the input-decoupling zeros for any input $u(t)$ and zero initial conditions.

Proof is similar to the proof of Theorem 4.1.

Example 6.1. (continuation of Example 5.2) In Example 5.2 it was shown that for the unreachable pair $(\overline{A}_2, \overline{B}_2 = 0)$ the matrix $\overline{A}_2$ has the form $\overline{A}_2 = \left[-\frac{R_2}{L_2} \right]$. Therefore, by Definition 6.1 the electrical circuit shown in Figure 5.3 has one input-decoupling zero $s_{i1} = -\frac{R_2}{L_2}$. Note that the input-decoupling zero corresponds to the mesh without the source voltage ($e_2 = 0$).

Definition 6.2. Let $\hat{A}_2$ be the matrix of unobservable part of the positive electrical circuit (2.3). The zeros $s_{o1}, s_{o2},..., s_{o\hat{n}_2}$ of the characteristic polynomial

$$\det[I_{\hat{n}_2} s - \hat{A}_2] = s^{\hat{n}_2} + \hat{a}_{\hat{n}_2 - 1} s^{\hat{n}_2 - 1} + ... + \hat{a}_1 s + \hat{a}_0 \tag{6.2}$$

of the matrix $\hat{A}_2$ are called the output-decoupling zeros of the positive electrical circuit (2.3).

The list of the output-decoupling zeros will be denoted by $Z_o = \{s_{o1}, s_{o2},..., s_{o\hat{n}_2}\}$.

Theorem 6.2. The output vector $y(t)$ of the positive electrical circuit is independent of the output-decoupling zeros for any input $\overline{u}(t) = Bu(t)$ and zero initial conditions.

Proof is similar to the proof of Theorem 4.1.

It is easy to show for the positive electrical circuit with $C = [0 \quad 0 \quad R_3]$ from Fig. 5.3 that the matrix $\hat{A}_2$ of the unobservable pair has the form

$$\hat{A}_2 = \begin{bmatrix} -\dfrac{R_1}{L_1} & 0 \\ 0 & -\dfrac{R_2}{L_2} \end{bmatrix}. \tag{6.3}$$

Therefore, by Definition 6.2 the positive electrical circuit shown in Fig. 5.2 has two output-decoupling zero $s_{o1} = -\dfrac{R_1}{L_1}$, $s_{o2} = -\dfrac{R_2}{L_2}$.

Definition 6.3. Zeros $s_{io}^{(1)}, s_{io}^{(2)}, ..., s_{io}^{(k)}$ which are simultaneously the input-decoupling zeros and the output-decoupling zeros of the positive electrical circuit are called the input-output decoupling zeros of the positive electrical circuit, i.e.

$$s_{io}^{(j)} \in Z_i \text{ and } s_{io}^{(j)} \in Z_o \text{ for } j = 1,2,\ldots,k;\ k \le \min(\bar{n}_2, \hat{n}_2). \tag{6.4}$$

The list of input-output decoupling zeros will be denoted by $Z_{io} = \{z_{io}^{(1)}, z_{io}^{(2)}, ..., z_{io}^{(k)}\}$.

Example 6.2. Consider the positive electrical circuit shown in Fig. 5.3 with the matrices *A*, *B*, *C* given by (5.10b) and $C = [0 \quad 0 \quad R_3]$. In Example 6.1 it was shown that the positive electrical circuit has one input-decoupling zero $s_{i1} = -\dfrac{R_2}{L_2}$ and two output-decoupling zeros $s_{o1} = -\dfrac{R_1}{L_1}, s_{o2} = -\dfrac{R_2}{L_2}$. Therefore, by Definition 4.3 the positive electrical circuit has one input-output decoupling zero $s_{io}^{(1)} = -\dfrac{R_2}{L_2}$.

These considerations can be extended to fractional electrical circuits [8].

7 Concluding Remarks

New necessary and sufficient conditions for the reachability and observability of the positive linear electrical circuits have been established. The definitions of the input-decoupling zeros, output-decoupling zeros and input-output decoupling zeros of the positive systems and electrical circuits have been proposed. Some properties of the new decoupling zeros have been discussed. The considerations have been illustrated by numerical examples of positive electrical circuits (systems) composed of resistors, coils and voltage source. An open problem is an extension of these considerations to fractional discrete-time and continuous-time positive linear systems and fractional electrical circuits [11].

Acknowledgment. This work was supported under work S/WE/1/11.

References

1. Antsaklis, P.J., Michel, A.N.: Linear Systems. Birkhauser, Boston (2006)
2. Farina, L., Rinaldi, S.: Positive Linear Systems; Theory and Applications. J. Wiley, New York (2000)
3. Kaczorek, T.: Positive 1D and 2D systems. Springer, London (2001)
4. Kaczorek, T.: Linear Control Systems, vol. 1. J. Wiley, New York (1993)
5. Kaczorek, T.: Decomposition of the pairs (A,B) and (A,C) of the positive discrete-time linear systems. Archives of Control Sciences 20(3), 341–361 (2010)
6. Kaczorek, T.: Controllability and observability of linear electrical circuits. Electrical Review 87(9a), 248–254 (2011)
7. Kaczorek, T.: Decoupling zeros of positive discrete-time linear systems. Circuit and Systems 1, 41–48 (2010)
8. Kaczorek, T.: Positivity and reachability of fractional electrical circuits. Acta Mechanica et Automatica 5(2), 42–51 (2011)
9. Kaczorek, T.: Reachability and controllability to zero tests for standard and positive fractional discrete-time systems. Journal Europeen des Systemes Automatises, JESA 42(6-8), 770–781 (2008)
10. Kaczorek, T.: Positive electrical circuits and their reachability. Archives of Electrical Engineering 60(3), 283–301 (2011)
11. Kaczorek, T.: Selected Problems of Fractional Systems Theory. Springer, Berlin (2011)
12. Kaczorek, T.: Decoupling zeros of positive continuous-time linear systems. Bull. Pol. Acad. Sci. Tech. 61(3) (2013)
13. Kaczorek, T.: Decoupling zeros of positive electrical circuits. Archives of Electrical Engineering 62(2) (2013)
14. Kailath, T.: Linear Systems. Prentice-Hall, Englewood Cliffs (1980)
15. Kalman, R.E.: Mathematical Descriptions of Linear Systems. SIAM J. Control 1, 152–192 (1963)
16. Kalman, R.E.: On the General Theory of Control Systems. In: Proc. of the First Intern. Congress on Automatic Control, pp. 481–493, Butterworth, London (1960)
17. Rosenbrock, H.H.: State-Space and Multivariable Theory. J. Wiley, New York (1970)
18. Tokarzewski, J.: Finite zeros of positive linear discrete-time systems. Bull. Pol. Acad. Sci. Tech. 59(3), 287–292 (2011)
19. Tokarzewski, J.: Finite zeros of positive continuous-time systems. Bull. Pol. Acad. Sci. Tech. 59(3), 293–302 (2011)
20. Tokarzewski, J.: Finite Zeros in Discrete-Time Control Systems. Springer, Berlin (2006)
21. Wolovich, W.A.: Linear Multivariable Systems. Springer, New York (1974)

Information Security of Nuclear Systems

Jason T. Harris

Idaho State University, Department of Nuclear Engineering and Health Physics, Idaho, USA
harrjaso@isu.edu

Abstract. The emphasis on information or cyber security has increased drastically over the last several years due to the increased number in attacks at the personal, company and even state level. Billions of Euros have been lost due to these attacks and the amount of funding expended to prevent them is even greater. One particular area of concern is in the protection of nuclear system information assets. These assets pertain to both information technology (IT) and instrumentation and control (I&C). Nuclear power plants (NPPs) are especially concerned as they transition from analog to digital technology. Nuclear information security garnered global attention at the recent 2012 Seoul Nuclear Security Summit and the 2013 International Conference on Nuclear Security: Enhancing Global Efforts. This paper discusses the information security domains at NPPs, the nuclear IT and I&C assets, and what these facilities are doing to protect themselves from cyber threats.

Keywords: Nuclear security, information security, nuclear power plant.

1 Introduction

The need for information security is not a new phenomenon. Encrypted messages have been used for thousands of years in areas such as warfare and politics. Cryptography, which is the study of techniques for secure communication, can be traced back to ancient Egypt and Greece. One of the earliest methods of encryption utilized a transposition cipher. The cipher contains a message in which the positions held by units of text are shifted according to a regular system. The cipher text, which has encoded or encrypted information, constitutes a permutation of the text and the order of units is changed. Some mathematical function is used on the characters' positions to encrypt, with an inverse function used to decrypt.

Since ancient times, many things have changed in regard to information security. First, both the speed at which information is processed and the throughput has increased significantly. There is also increased reliance on information and the information itself can be much more complex. There is an increased use of computerized, networked control systems and a human mediator may not filter information. Finally there are increased risks related to knowledge of information.

Information security faces new challenges in today's world due to the emergence of hacking, cyberterrorism, and cyberwarfare. One very specific area of information security where there is increasing concern is in nuclear systems. Information security

J. Świątek et al. (eds.), *Advances in Systems Science*,
Advances in Intelligent Systems and Computing 240,
DOI: 10.1007/978-3-319-01857-7_2,

is an integral part of nuclear security. Nuclear information security is concerned with the protection of information assets from a wide range of threats specific to nuclear systems. The objectives of nuclear information security include: protection against loss of nuclear sensitive and/or classified information; protection against the theft of material (both physical and information); protection against terrorist actions and sabotage; protections against a combined cyber and physical attack, ensuring nuclear safety, and ensuring business continuity [3]. Consequences of an event related to nuclear information can affect the owner of the information or system, organizations and individuals responsible for secure and safe operation of the process, the government, and possibly the public.

The importance of nuclear information security was made clear at two recent and significant conferences: the Nuclear Security Summit held in Seoul, Korea in 2012 and the International Conference on Nuclear Security held in Vienna in 2013. The 2012 Nuclear Security Summit had participation from more than 53 heads of state and international organizations. The International Conference on Nuclear Security held by the International Atomic Energy Agency (IAEA) had participations from nearly 600 governmental representatives and nuclear security technical experts from all over the world. Both meetings focused on the importance of preventing theft of information, technology or expertise required to acquire and use nuclear materials for malicious purposes, or to disrupt information technology based control systems at nuclear facilities [1, 2]. The loss of nuclear information at Los Alamos National Laboratory, and recent cyber attacks on the Davis-Besse nuclear power plant and Iranian nuclear facilities are proof that the threats are real.

2 Security Concepts

Before discussing nuclear information security, it is necessary to describe basic security attributes and information assets. The three basic attributes of all security-related systems are confidentiality, integrity, and availability (CIA approach). Confidentiality means ensuring that unauthorized people, resources, or processes cannot access information. Integrity involves the protection of information from intentional or accidental unauthorized changes. Availability is assurance that information is available whenever needed. Both integrity and availability are critical for real-time control applications. For nuclear facilities, two additional security principles can be applied and are important when safety needs to be considered. These facilities need to be free from interference and be robust against undesired attacks.

Information assets are all elements of information that either shares a common usage, purpose, associated risk and/or form of storage. These assets are also defined as sets of information that are considers of value to an individual, organization, or State. Information security is the regime or program in place to ensure the protection of these assets.

2.1 Information Security Objectives

Information, or more specifically computer, security requires protecting the confidentiality, integrity and availability attributes of electronic data or computer

systems and processes. By identifying and protecting these attributes in data or systems that can have an adverse impact on the safety and security functions in facilities (nuclear), the security objectives can be met [4]. Computer security, used in this context, is the same as information technology (IT) security or cyber security.

Security of computer systems, regardless of location, requires specific measures. Preventative measures are set up to protect against threats. Measures and processes must also be in place to detect threats, both single and continuous. Processes and instructions must also be in place to properly respond or react to threats or attacks. Measures must also include an evaluation of risk. Risk is the potential that a given threat will exploit the vulnerabilities of an asset or group of assets. It is measured in terms of a combination of the likelihood of an event and the severity of its consequences.

2.2 Nuclear Security Objectives

Information security at nuclear facilities is a crucial component of an overall nuclear security plan. This nuclear security plan is developed as part of a State's nuclear security regime. The International Atomic Energy Agency (IAEA) has developed nuclear security recommendations on radioactive material and associated facilities. The overall objective of a State's nuclear security regime is to protect persons, property, society, and the environment from malicious acts involving nuclear material or other radioactive material that could cause unacceptable radiological consequences. Ultimately, protection should be from both unauthorized removal (theft) of material and acts of sabotage [5].

Nuclear security relies on the identification and assessment of those assets that need to be secured. Classification of sensitive information assets defines how they should be protected. In a nuclear facility, there are three types of computer (digital) equipment assets. Information technology (IT) consists of computers, networks and databases. Industrial control systems (ICS) encompass several types of control systems used in industrial production and critical infrastructures. This includes supervisory control and data acquisition (SCADA) systems and other control system configurations such as programmable logic controllers (PLC). Instrumentation and control (I&C) systems support nuclear production processes in plants. The IAEA has determined there are three groups of threats on computer systems at nuclear facilities: information gathering attacks aimed at planning and executing further malicious attacks; attacks disabling or compromising attributes of one or several computers crucial to facility security or safety; and attacks on computers combined with a physical intrusion to target locations [4].

Implementing adequate security and countermeasures at nuclear facilities is very different than in other business sectors. The typical approach for protection of a company's assets depends on risk appetite and tolerances. Unrestricted use of this approach is not acceptable for nuclear systems. Security risks are partially determined by a State. The State develops a Design Basis Threat (DBT) that outlines security requirements and then is given to the facility. The DBT takes into account all

credible, postulated, and perceived threats that various assets within a State may be vulnerable to a threat or attack. These threats may be physical (i.e. sabotage, explosion, theft) or cyber in nature. The DBT is developed by the State with input from various governmental agencies, regulators, and facilities.

3 Nuclear Information Security Domain

Operations at nuclear facilities are very unique and as such, the computer security requirements are equally unique. Nuclear operations can include several modes such as start-up, operating power, hot shutdown, cold shutdown, and refueling. Testing of systems is also done at various time intervals (i.e. daily, weekly, and monthly). The need for guidance addressing computer security at nuclear facilities is supported by the special conditions characterizing the industry. Nuclear facilities must abide by requirements set by their national regulatory bodies, which may directly or indirectly regulate computer systems or set guidance. Nuclear facilities may have to protect against additional threats, which are not commonly considered in other industries. Such threats may also be induced by the sensitive nature of the nuclear industry. Computer security requirements in nuclear facilities may differ from requirements in other concerns. Typical business operations involve only a limited range of requirements. Nuclear facilities need to take a wider base or an entirely different set of considerations into account.

Nuclear information security domains vary depending on the type of facility. This paper will not cover the multitude of nuclear facilities (enrichment facilities, fuel fabrication facilities, waste storage facilities, etc) available. Instead, the nuclear information security of a nuclear power plant (NPP) will be expounded upon. Nuclear power plants are found in over 30 countries and represent one of the most credible sources for a cyber attack. The information security domain of a typical NPP is shown in Figure 1 [6].

Information and control (I&C) assets are all digital elements that are used for safety functions or operational control, especially for measurement, actuation, and monitoring plant parameters. Examples of these assets include reactor protection systems, emergency core cooling, emergency power supply, reactor power control, fire detection, and radiation monitoring. General components of I&C systems include micro-controllers, drivers, motors, valves, servers, and gateways. Most information security processes are focused on IT environments. However, great changes to digital based technologies on I&C systems are occurring in nuclear facilities. Previously control systems associated with nuclear power plants existed as isolated “islands”, but that has changed. In a NPP control room, the operator workstation will be connected by a number of operations, maintenance, data acquisition, and process handling data highways. Control and protection related systems are independent of these data highways. Vital plant control systems will utilize redundant communications. Protection systems will utilize redundant channels.

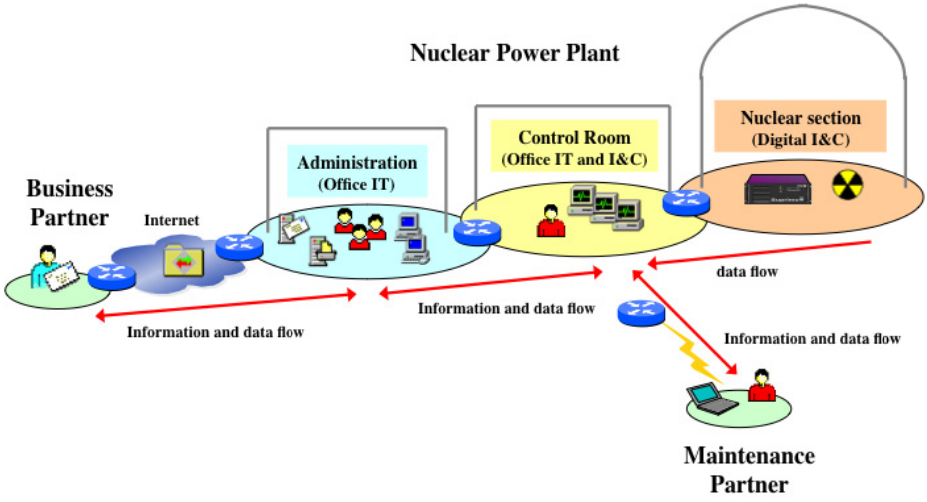

Fig. 1. Information security domains at a nuclear power plant (Source: INSEN, p. 7)

The importance of I&C asset security cannot be overstated. But information security at a NPP is equally important. At the core of information security is the development of an information security management system (ISMS). The ISMS means: understanding an organization's information security requirements and the need to establish policy and objectives for information security; implementing and operating controls to manage an organization's information security risks in the context of the organization's overall business risks; monitoring and reviewing the performance and effectiveness of the ISMS; and continual improvement based on objective measurement. At a NPP, the specific drivers to the ISMS process include generally increased IT security requirements (critical infrastructure), legal requirements, directives and guidelines, co-determination by authorities, and consideration of insider threat. Subject areas focus on management, business processes, the plant, operations, information systems, access control, and the workforce.

4 Protecting Nuclear Information Assets

Protecting information assets requires the integration of several approaches. First, a holistic approach must be taken. Implementation of this approach requires identifying and classifying all information assets, assessing threats, vulnerabilities and risk, and applying graded protective measures according to the asset's life cycle. All disciplines of security at the facility must interact and complement each other to establish a facility's security posture. Next, human factors must be considered. Key to this is the development of a strong security culture. Security culture is a set of beliefs, attitudes, behaviors, and management systems that lead to a more effective nuclear security program. This includes information /computer security culture. As humans are typically the weakest link when it comes to security, ongoing training and strong management is very important. Heavy security in nuclear facilities must also be balanced with acceptance by the workforce.

There are a number of concepts that are utilized to protect nuclear information. Access control, authorization, and need-to-know concepts are used extensively in a variety of business environments. Nuclear facilities also incorporate a graded approach to security as well as defense-in-depth. The graded approach uses security measures that are adequate to the protection level needed. As the criticality of the systems increases, so do the level of security/strength of measures. The defense-in-depth principle arranges information assets in a way that low sensitive assets are easier to access than high sensitive assets (i.e. zone model with subordinate barriers). In a NPP, there will be internet and network zone borders. Technical concepts focus on the protection of the information assets itself and of the infrastructure the information assets use. Technical measures impact organizational issues like IT processes. Examples of technical concepts include network security, authentication and crypotography, intrusion detection, and network management. Finally, organizational concepts focus on the processes dealing with the information assets. Security must be implemented in the processes that are usually driven by the IT department. Subject areas include purchasing, software development, problem management, incident management, and help desk.

Nuclear facilities have a number of special considerations that need to be addressed when it comes to information security. First, facility lifetime phases and modes of operation must be evaluated. Access of information may vary drastically at these different times. Also there will be security requirement differences between IT and I&C systems. The potential consequences that come with the demand for additional connectivity must be carefully considered. Software updates, secure design, and specifications for computer systems throughput the facility must be evaluated. Finally, third party or vendor access control procedures must be developed, evaluated, and updated on a continuous basis.

5 Conclusions

As the number of cyber and information attacks increases, facilities must be increasingly vigilant of protecting their assets. This is especially true for nuclear facilities where compromised security can lead to degradation of safety systems, which in turn can lead to detrimental consequences to the facility, humans, and the environment. It is of utmost importance that information security professionals are trained properly and can stay ahead of the threat. Educational initiatives, much like what has been developed by the IAEA, are crucial for developing a workforce that can keep nuclear facilities safe and secure from malicious acts.

References

1. Nuclear Security Summit: Seoul Communiqué. Nuclear Security Summit. IAEA, Vienna (March 27, 2012)
2. International Atomic Energy Agency: Proceedings of the International Conference on Nuclear Security: Enhancing Global Efforts. IAEA, Vienna (2013)

3. International Atomic Energy Agency: Introduction to Information and Computer Security. IAEA Introduction to Nuclear Security Module 14. IAEA, Vienna (2012)
4. International Atomic Energy Agency: Computer Security at Nuclear Facilities. IAEA Nuclear Security Series No. 17. IAEA, Vienna (2011)
5. International Atomic Energy Agency: Nuclear Security Recommendations on Radioactive Material and Associated Facilities. IAEA Nuclear Security Series No. 14. IAEA, Vienna (2011)
6. International Atomic Energy Agency: NS 1 Introduction to Nuclear Security: NS 1.11 Information Security. INSEN Educational Material, pp. 1–60. IAEA, Vienna (2012)
7. International Atomic Energy Agency: Core Knowledge on Instrumentation and Control Systems in Nuclear Power Plants. IAEA Nuclear Energy Series No. NP-T-3.12. IAEA, Vienna (2011)

Distributed Reconfigurable Predictive Control of a Water Delivery Canal

João M. Lemos[1], José M. Igreja[2], and Inês Sampaio[1]

[1] INESC-ID/IST, Technical University of Lisbon, Lisboa, Portugal
jlml@inesc-id.pt
[2] INESC-ID/ISEL, Lisboa, Portugal
jigreja@deea.isel.ipl.pt

Abstract. This paper addresses the problem of reconfigurable distributed model predictive control (MPC) of water delivery canals. It is shown how a distributed MPC algorithm can be equipped with complementary features so as to reconfigure its structure in order to render it tolerant to actuator faults. The structure proposed includes a fault detection algorithm that triggers switching between different controllers designed to match the fault or no-fault situation. To ensure stability, a dwell-time switching logic is used. Experimental results are provided.

Keywords: Fault tolerant control, reconfigurable control, distributed control, predictive control, water delivery canal.

1 Introduction

Water delivery open canals used for irrigation [1] are large structures whose complexity, together with increasing requirements on reliability and quality of service provides a strong motivation to consider fault tolerant control methods [2]. In order to achieve fault tolerant features, the idea consists in exploring the redundancy in instaled sensors and actuators to reconfigure the control system such as to allow the plant operation to continue, perhaps with some graceful degradation, when a sensor or actuator fails.

The concept of fault tolerant control (FTC) has been the subject of intense research in the last twenty years [3–5], in particular in what concerns reconfigurable fault tolerant control systems [7]. This activity yielded a rich bibliography that, of course, cannot be covered here and that comprises aspects such as fault detection and isolation and fault tolerant control design. In relation to distributed control, an important concept is "integrity", namely the capacity of the system to continue in operation when some part of it fails [6]. Other type of approach models the failures as disturbances that are estimated and compensated by the controller [8]. In what concerns water delivery canal systems topics found in the literature include control loop monitoring [13] and reconfiguration to mitigate fault effects [12]. Reconfiguring the controller in face of a plant fault falls in the realm of hybrid systems and raises issues related to stability that must be taken into account [10].

J. Świątek et al. (eds.), *Advances in Systems Science*,
Advances in Intelligent Systems and Computing 240,
DOI: 10.1007/978-3-319-01857-7_3,

The contribution of this paper consists of the application of MPC based distributed fault tolerant control to a water delivery canal in the presence of actuator faults. An algorithm based on controller reconfiguration with a dwell time logic is presented, together with experimental results.

The paper is organized as follows: After the introduction in which the work is motivated, a short literature review is made and the main contributions are presented, the canal is described in section 2, including a static nonlinearity compensation of the gate model. Distributed MPC control is described in section 3, whereas actuator fault tolerant control is dwelt with in section 4. Experimental results are presented in section 5. Finally, section 6 draws conclusions.

2 The Canal System

2.1 Canal Description

The experimental work reported hereafter was performed at the large scale pilot canal of *Núcleo de Hidráulica e Controlo de Canais* (Universidade de Évora, Portugal), described in [11]. The canal has four pools with a length of 35m, separated by three undershoot gates, with the last pool ended by an overshoot gate. In this work, only the first three gates are used. The maximum nominal design flow is $0.09\,\mathrm{m^3 s^{-1}}$. There are water off-takes downstream from each branch made of orifices in the canal walls, that are used to generate disturbances corresponding to water usage.

Water level sensors are installed downstream of each pool. The water level sensors allow to measure values between 0 mm and 900 mm, a value that corresponds to the canal bank. For pool number i, $i = 1, \ldots, 4$, the downstream level is denoted y_i and the opening of gate i is denoted u_i. The nomenclature is such that pool number i ends with gate number i. Each of the actual gate positions $u_{r,i}$, $i = 1, 2, 3$ is manipulated by a command signal u_i.

2.2 Nonlinearity Compensation

Following [2], in order to compensate for a nonlinearity, instead of using as manipulated variable the gate positions $u_{r,i}$, the corresponding water flows q_i crossing the gates are used. These are related by

$$q_i = C_{ds} W u_{r,i} \sqrt{2g(h_{upstr,i} - h_{downstr,i})}, \tag{1}$$

where C_{ds} is the discharge coefficient, W is the gate width, $g = 9,8\mathrm{m/s}$ is the gravity acceleration, $h_{upstr,i}$ is the water level immediately upstream of the gate and $h_{downstr,i}$ is the water level immediately downstream of the gate. This approach corresponds to representing the canal by a Hammerstein model and to compensating the input nonlinearity using its inverse. The linear controller computes the flow crossing the gates, that is considered to be a virtual command variable v_i and the corresponding gate position is then computed using (1). The discharge coefficient is not estimated separately, but instead is considered to be incorporated in the static gain of the linear plant model.

2.3 Canal Model

In order to design the controllers, the dynamics of the canal has been approximated by a finite dimension linear state-space model written as

$$x(k+1) = Ax(k) + Bv(k), \tag{2}$$

$$y(k) = Cx(k) \tag{3}$$

where $k \in \mathbb{N}$ denotes discrete time, $x \in \mathbb{R}^n$ is the full canal state, $y \in \mathbb{R}^p$ is the output made of the downstream pool levels, with $p = 3$ the number of outputs, $v \in \mathbb{R}^p$ is the manipulated variable and $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$ and $C \in \mathbb{R}^{p\times n}$ are matrices. Assuming operation around a constant equilibrium point, these matrices are identified by constraining the model to have the following structure

$$A = \begin{bmatrix} A_{11} & \underline{0} & \underline{0} \\ \underline{0} & A_{22} & \underline{0} \\ \underline{0} & \underline{0} & A_{33} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} & \underline{0} \\ B_{21} & B_{22} & B_{23} \\ \underline{0} & B_{32} & B_{33} \end{bmatrix}, \quad C = \begin{bmatrix} C_1 & \underline{0} & \underline{0} \\ \underline{0} & C_2 & \underline{0} \\ \underline{0} & \underline{0} & C_3 \end{bmatrix}. \tag{4}$$

These matrices have dimensions that match the state x_i associated to each pool such that $x = [x_1' x_2' x_3']'$. This structure is imposed to reflect the decomposition of the canal model in subsystems, each associated to a different pool. Furthermore, it is assumed that each pool interacts directly only with its neighbors, and only through the input.

3 Distributed Siorhc

3.1 A Strategy for Distributed Control

Let the canal be decomposed in a number of local linear time invariant subsystems Σ_i, $i = 1, \ldots, N_\Sigma$. Each of these subsystems have a manipulated input u_i and a measured output y_i. In addition, Σ_i interacts with its neighbors, Σ_{i-1} and Σ_{i+1}. It is assumed that this interaction takes only part through the manipulated variables and that the cross-coupling of states can be neglected.

As shown in figure 1, a local controller $\mathscr{C}_i$ is associated to each subsystem Σ_i. At the beginning of each sampling interval, this local controller computes the value of the manipulated variable using knowledge of y_i but also by performing a negotiation with the adjacent controllers $\mathscr{C}_{i-1}$ and $\mathscr{C}_{i+1}$. This negotiation takes place in a recursive way along the following steps:

1. Let l be the step index and $\Delta U_i(k, l)$ be the increment of the manipulated variable of local controller i at sampling time k and after performing l steps.
2. Set the counter $l = 1$.
3. Assume that each local controller i, $i = 1, 2, 3$, at time k and after performing l steps knows $\Delta U_{i-1}(k, l-1)$ and $\Delta U_{i+1}(k, l-1)$, *i. e.*, each local controller knows the previous iteration of the neighbor controllers. Update the control increment of each local controller by

$$\Delta U_i(k, l) = \mathscr{F}(\Delta U_{i-1}(k, l-1), \Delta U_{i+1}(k, l-1)) \tag{5}$$

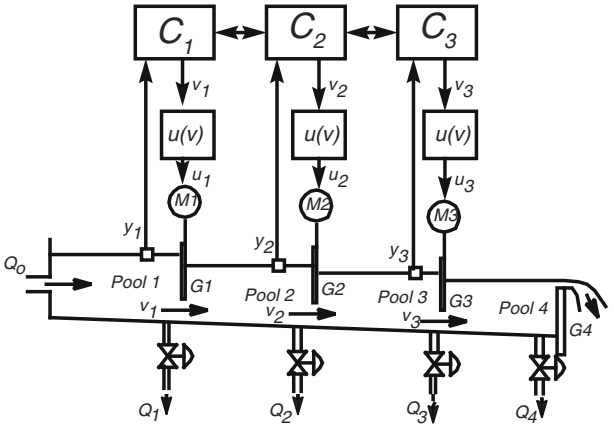

Fig. 1. Distributed controller in normal (no fault) operation

where $\mathscr{F}$ denotes the optimization procedure used, that varies from algorithm to algorithm.

4. If convergence is reached, stop. Otherwise, set $l \rightarrow l+1$ and go to step 2.

In the next sub-section, a distributed version of a model predictive controller with stability constraints, named SIORHC, is obtained using this procedure.

3.2 Distributed Siorhc

Consider the system described by the linear state model (2), augmented with an integrator. For this system, the predicted outputs at $k+j$ given observations up to time k are given by:

$$\begin{aligned} \hat{y}(k+j) &= \sum_{i=0}^{j-1} C\ A^{j-i-1} B \Delta u(k+i) + \hat{y}_0(k+j) \\ \hat{y}_0(k+j) &= C\ A_j\ \hat{x}(k) \end{aligned} \tag{6}$$

where $\hat{y}_0$ is the output predict value without control moves (the system free response) and $\hat{x}$ denote either the state or its estimate obtained with a suitable observer. For $j = 1...N,\ N+1, ...,\ N+P$ (6) the predictors can be written in a compact way as

$$\begin{aligned} \hat{Y}_N &= G_N \Delta U + \hat{Y}_{0N} \\ \hat{Y}_P &= G_P \Delta U + \hat{Y}_{0P} \end{aligned} \tag{7}$$

with

$$\begin{aligned} \hat{Y}_N &= [y_{k+1} \cdots y_{k+N}]^T \\ \hat{Y}_P &= [y_{k+N+1} \cdots y_{k+N+P}]^T \end{aligned} \tag{8}$$

In order to develop a distributed controller version, let the system be decomposed in a number of serially connected subsystems Σ_i, $i = 1, \ldots, N_\Sigma$. For the sake of clarity consider the case $N_\Sigma = 3$. Equation (7) is then approximated considering only interactions between neighboring serially connected systems:

$$\begin{aligned}
\hat{Y}_{1N} &= \hat{Y}_{10N} + G_{11N}\Delta U_1 + G_{12N}\Delta U_2 \\
\hat{Y}_{1P} &= \hat{Y}_{10P} + G_{11P}\Delta U_1 + G_{12P}\Delta U_2 \\
\hat{Y}_{2N} &= \hat{Y}_{20N} + G_{21N}\Delta U_1 + G_{22N}\Delta U_2 + G_{23N}\Delta U_3 \\
\hat{Y}_{2P} &= \hat{Y}_{20P} + G_{21P}\Delta U_1 + G_{22P}\Delta U_2 + G_{23P}\Delta U_3 \\
\hat{Y}_{3N} &= \hat{Y}_{30N} + G_{32N}\Delta U_2 + G_{33N}\Delta U_3 \\
\hat{Y}_{3P} &= \hat{Y}_{30P} + G_{32P}\Delta U_2 + G_{33P}\Delta U_3
\end{aligned}$$

Associate the following local cost functional to Σ_1:

$$\overline{J_1} = \sum_{i=1}^{N} e_{1,\,k+i}^T Q_1 e_{1,\,k+i} + \sum_{i=1}^{N} e_{2,\,k+i}^T Q_2 e_{2,\,k+i} + \sum_{i=0}^{N-1} \Delta u_{1,\,k+i}^T R_1 \Delta u_{1,\,k+i} \quad (9)$$

with zero terminal horizon constraint given by:

$$[y_{1,\,k+N+1} \;\cdots\; y_{1,\,k+N+P}]^T = [r_{1,\,k+N+1} \;\cdots\; r_{1,\,k+N+P}]^T \quad (10)$$

where $e_{(.),k} = r_{(.),k} - y_{(.),k}$, is the tracking error in relation to the reference sequence, $r_{(.),k}$, and $Q_{(.)} \geq 0$ and $R_{(.)} > 0$ are weighting matrices.

In an equivalent way, the minimization of $\overline{J_1}$ with respect to ΔU_1 may be written as:

$$\min_{\Delta U_1} \overline{J_1} = \|Y_{1RN} - \hat{Y}_{1N}\|_{Q_1}^2 + \|Y_{2RN} - \hat{Y}_{2N}\|_{Q_2}^2 + \|\Delta U_1\|_{R_1}^2 \quad (11)$$

$$s.t. \quad \hat{Y}_{1P} = Y_{1RP}$$

The stated QP optimization problem with constraints can now be solved by finding the vector ΔU_1 that minimizes the Lagrangian:

$$\begin{aligned}
\mathscr{L}_1 := {} & \|E_1 - G_{11N}\Delta U_1 - G_{12N}\Delta U_2\|_{Q1}^2 + \\
& + \|E_2 - G_{21N}\Delta U_1 - G_{22N}\Delta U_2 - G_{23N}\Delta U_3\|_{Q2}^2 + \\
& + \|\Delta U_1\|_{R1}^2 + \|\Delta U_2\|_{R2}^2 + \\
& + [F_1 + G_{11P}\Delta U_1 + G_{12P}\Delta U_2]^T \lambda_1
\end{aligned} \quad (12)$$

where $E_j = Y_{jRN} - \hat{Y}_{j0N}$, $F_j = Y_{jRP} - \hat{Y}_{j0P}$ and λ_1 is a column vector of Lagrange multipliers. Solving (12) yields

$$M_1 \Delta U_1 = \left(I - G_{11P}^T W_1 G_{11P} M_1^{-1}\right)\left(G_{11N}^T Q_1 E_1 + G_{21N}^T Q_2 E_2\right) + W_1 F_1 \quad (13)$$

with

$$W_1 = \left(G_{11P} M_1^{-1} G_{11P}^T\right)^{-1} \quad (14)$$

$$M_1 = G_{11N}^T Q_1 G_{11N} + G_{21N}^T Q_2 G_{11N} + R_1 \quad (15)$$

Using analogous procedures, another two equations are obtained for the controllers associated with Σ_2 and Σ_3 by minimizing the local functionals:

$$\min_{\Delta U_2} \overline{J_2} = \|Y_{1RN} - \hat{Y}_{1N}\|^2_{Q_1} + \|Y_{2RN} - \hat{Y}_{2N}\|^2_{Q_2} + \|Y_{3RN} - \hat{Y}_{3N}\|^2_{Q_3} + \|\Delta U_2\|^2_{R_2} \quad (16)$$

$$s.t. \quad \hat{Y}_{2P} = Y_{2RP}$$

and

$$\min_{\Delta U_3} \overline{J_3} = \|Y_{2RN} - \hat{Y}_{2N}\|^2_{Q_2} + \|Y_{3RN} - \hat{Y}_{3N}\|^2_{Q_3} + \|\Delta U_3\|^2_{R_3} \quad (17)$$

$$s.t. \hat{Y}_{3P} = Y_{3RP}$$

yielding

$$M_2 \Delta U_2 = \left(I - G^T_{22P} W_2 G_{22P} M_2^{-1}\right) \left(G^T_{12N} Q_1 E_1 + \right.$$
$$\left. + G^T_{22N} Q_2 E_2 + G^T_{32N} Q_3 E_3\right) + G^T_{22P} W_2 F_2 \quad (18)$$

and

$$M_3 \Delta U_3 = \left(I - G^T_{33P} W_3 G_{33P} M_3^{-1}\right) \left(G^T_{23N} Q_2 E_2 + \right.$$
$$\left. + G^T_{33N} Q_3 E_3\right) + G^T_{33P} W_3 F_3 \quad (19)$$

The distributed SIORHC solution for the serially connected sub-systems can be obtained, using the procedure in subsection 3.1, from the matrix algebraic equations system:

$$\Phi \Delta U = \Psi \quad (20)$$

where the Φ matrix building blocks are:

$$\begin{aligned}
\Phi_{11} &= M_1 \\
\Phi_{12} &= S_1 \left(G^T_{11N} Q_1 G_{12N} + G^T_{21N} Q_2 G_{22N}\right) + G^T_{11P} W_1 G_{12P} \\
\Phi_{13} &= S_1 \left(G^T_{21N} Q_2 G_{23N}\right) \\
\Phi_{21} &= S_2 \left(G^T_{12N} Q_1 G_{11N} + G^T_{22N} Q_2 G_{21N}\right) + G^T_{22P} W_2 G_{21P} \\
\Phi_{22} &= M_2 \\
\Phi_{23} &= S_2 \left(G^T_{22N} Q_2 G_{23N} + G^T_{32N} Q_3 G_{33N}\right) + G^T_{22P} W_2 G_{23P} \\
\Phi_{31} &= S_3 \left(G^T_{23N} Q_2 G_{21N}\right) \\
\Phi_{32} &= S_3 \left(G^T_{23N} Q_2 G_{22N} + G^T_{33N} Q_3 G_{32N}\right) + G^T_{33P} W_3 G_{32P} \\
\Phi_{33} &= M_3
\end{aligned} \quad (21)$$

with $S_i = I - G^T_{iiP} W_i G_{iiP} M_i^{-1}$, $i = 1, 2, 3$, the entries of Ψ

$$\begin{aligned}
\Psi_1 &= S_1 \left(G^T_{11N} Q_1 A_1 + G^T_{21N} Q_2 A_2\right) + G^T_{11P} W_1 B_1 \\
\Psi_2 &= S_2 \left(G^T_{12N} Q_1 A_1 + G^T_{22N} Q_2 A_2 + G^T_{32N} Q_3 A_3\right) + G^T_{22P} W_2 B_2 \\
\Psi_3 &= S_3 \left(G^T_{23N} Q_2 A_2 + G^T_{33N} Q_3 A_3\right) + G^T_{33P} W_3 B_3
\end{aligned} \quad (22)$$

and

$$\Delta U = \left[\, \Delta U_1 \; \Delta U_2 \; \Delta U_3 \,\right] \quad (23)$$

To apply the iterative procedure described in subsection 3.1, write (20) as

$$\Phi_d \Delta U(k,\, l+1) + \Phi_{nd} \Delta U(k,\, l) = \Psi \tag{24}$$

where

$$\Phi_d = \begin{bmatrix} \Phi_{11} & 0 & 0 \\ 0 & \Phi_{22} & 0 \\ 0 & 0 & \Phi_{11} \end{bmatrix} \quad \Phi_{nd} = \begin{bmatrix} 0 & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & 0 & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & 0 \end{bmatrix}. \tag{25}$$

The algorithm will converge provided that the spectral radius

$$\lambda_{max} := \max\, eig\left(\Phi_d^{-1} \Phi_{nd}\right) \tag{26}$$

verifies

$$|\lambda_{max}| < 1 \tag{27}$$

4 Fault Tolerant Control

4.1 Controller Reconfiguration

Figure 2 shows a discrete state diagram that explains how controller reconfiguration is performed when an actuator fault occurs in the water channel considered in this paper. For simplicity, only the occurrence of faults in gate 2 are considered. State $\mathcal{S}_1$ corresponds to the situation in which all gates are working normally with a controller $\mathcal{C}_N$ that matches this situation. When a fault occurs, the system state switches to $\mathcal{S}_2$, in which gate 2 is faulty (blocked) but the controller used is still the one designed for the no fault situation.

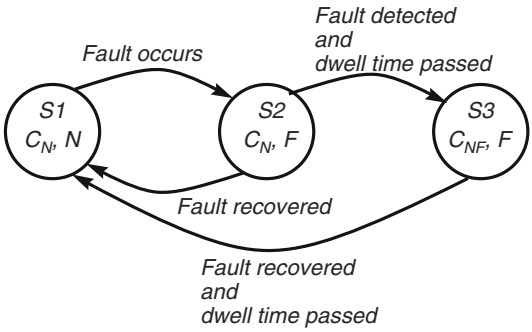

Fig. 2. Discrete states in controller reconfiguration

In the presence of a fault, the matrices of the state-space model (3) have the structure

$$A = \begin{bmatrix} A_{11}^F & \underline{0} \\ \underline{0} & A_{33}^F \end{bmatrix}, \quad B = \begin{bmatrix} B_{11}^F & B_{13}^F \\ B_{31}^F & B_{33}^F \end{bmatrix}, \quad C = \begin{bmatrix} C_1^F & \underline{0} \\ \underline{0} & C_3^F \end{bmatrix}. \tag{28}$$

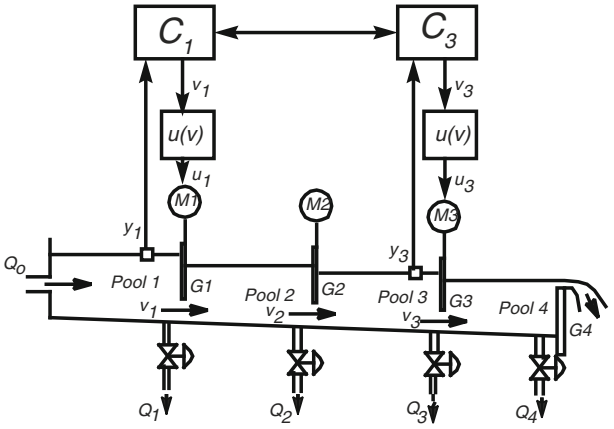

Fig. 3. Distributed controller structure after a fault is detected

The superscript F enhances the fact that the matrix blocks are estimated assuming that a fault has occurred and that they are different from the ones in (4). Figure 3 shows the controller to apply under a faulty situation. Controller reconfiguration implies a reconfiguration of the communication network as well.

When the fault is detected, the state switches to $\mathcal{S}_3$, in which a controller $\mathcal{C}_F$ designed for the faulty situation is connected to the canal. When the fault is recovered (gate 2 returns to normal operation), the state returns to $\mathcal{S}_1$. A dwell time condition is imposed to avoid instability that might arise due to fast switching [14]. This means that, once a controller is applied to the plant, it will remain so for at least a minimum time period (called dwell time). Furthermore, an integrator in series with the plant ensures bumpless transfer between controllers.

When distributed control is used, the controller designed for normal operation (shown in figure 1), $\mathcal{C}_N$, consists of 3 SISO SIORHC controllers $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$, each regulating a pool and such that each individual controller negotiates the control variable with its neighbors. This means that, in states $\mathcal{S}_1$ and $\mathcal{S}_2$, $\mathcal{C}_1$ negotiates with $\mathcal{C}_2$, $\mathcal{C}_2$ negotiates with $\mathcal{C}_1$ and with $\mathcal{C}_3$ and $\mathcal{C}_3$ negotiates with $\mathcal{C}_2$. The controller for the faulty condition (shown in figure 3) is made just of two SISO controllers that control pools 1 and 3 and negotiate with each other.

4.2 Fault Detection

For actuator faults, the fault detection algorithm operates as follows. For each gate i, $i = 1, 2, 3$, define the error $\tilde{u}_i$ between the command of the gate position u_i and the actual gate position $u_{r,i}$,

$$\tilde{u}_i(k) = u_i(k) - u_{r,i}(k) \tag{29}$$

A performance index Π is computed from this error by

$$\Pi(k) = \gamma\Pi(k-1) + (1-\gamma)|\tilde{u}(k)|. \tag{30}$$

If $\Pi(k) \geq \Pi_{max}$, where Π_{max} is a given threshold, then it is decided that a fault has occurred.

5 Experimental Results

Figure 4 show experimental results with D-SIORHC. At the time instant marked by a red vertical line, a fault that forces gate 2 to become stuck occurs. Shortly after, at the instant marked by the yellow vertical line, this fault is detected, and the controller is reconfigured as explained. From this moment on, there is no warranty on the value of the level J_2, but J_1 and J_3 continue to be controlled.

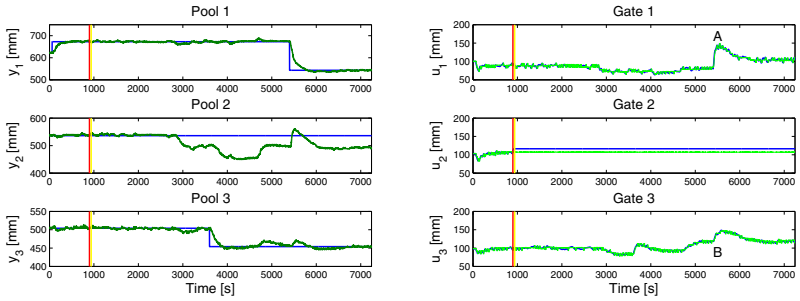

Fig. 4. D-SIORHC of the water delivery canal. Reconfiguration after a fault in gate 2.

This experiment also includes a test with respect to disturbance rejection. At time $t = 2800$ s, a disturbance is created by opening the side take of pool 1, thereby extracting some water. The controllers react by closing the gates in order to compensate for this loss of incoming flow.

6 Conclusions

A reconfigurable MPC controller with stability terminal constraints for a water delivery canal has been developed and demonstrated experimentally. Fault tolerance is embedded in the reconfiguration of a network of local controllers, that change their pattern of negotiation with neighbors in order to reach a consensus. Stability of this switched controller network is ensured by forcing a dwell time switching logic. It is possible to provide lower bounds on the dwell time that ensure stability of the overall system.

The experimental results presented illustrate that the system is able to tackle both the problems of disturbance rejection and reference tracking.

One difficulty stems from the fact that the PLCs that control gate motors impose a quantization effect. Therefore, the gate position changes only if the order for the new position differs from the actual position by at least 5 mm. This quantization effect imposes a limit on the performance of the overall system.

Acknowledgment. This work was supported by "FCT–Fundação para a Ciência e a Tecnologia", within projects *ORCHESTRA – Distributed Optimization and Control of Large Scale Water Delivery Systems* , contract PTDC/EMS-CRO/2042/2012, and PEst-OE/EEI/LA0021/2013.

References

1. Litrico, X., Fromion, V.: Modeling and control of hydrosystems. Springer (2009)
2. Cantoni, M., Weyer, E., Li, Y., Ooi, S.K., Mareels, I., Ryan, M.: Control of large-scale irrigation networks. Proc. IEEE 95(1), 75–91 (2007)
3. Blanke, M., Kinnaert, M., Lunze, J., Staroswiecki, M.: Diagnosis and Fault Tolerant Control, 2nd edn. Springer (2006)
4. Åstrom, K.J., Albertos, P., Blanke, M., Isidori, A., Schaufelberger, W., Sanz, R. (eds.): Control of complex systems. Springer (2001)
5. Blanke, M., Staroswiecki, M., Wu, N.E.: Concepts and methods in fault-tolerant control. In: Proc. 2001 Am. Control Conf., pp. 2606–2620 (2011)
6. Campo, P.J., Morari, M.: Achievable closed-loop properties of systems under decentralized control: Conditions involving the steady-state gain. IEEE Trans. Autom. Control 39(5), 932–943 (1994)
7. Zhang, Y., Jiang, J.: Bibliographical review on reconfigurable fault-tolerant control systems. Annual Reviews in Control 32, 229–252 (2008)
8. Zhao, Q., Jiang, J.: Reliable state feedback control design against actuator failures. Automatica 34(10), 1267–1272 (1998)
9. Weyer, E., Bastin, G.: Leak detection in open water channel. In: Proc. 17th IFAC World Congress, Seoul, Korea, pp. 7913–7918 (2008)
10. Koutsoukos, X.D., Antsaklis, P.J., Stiver, J.A., Lemmon, M.D.: Supervisory control of hybrid systems. Proc. of the IEEE 88(7), 1026–1049 (2000)
11. Lemos, J.M., Machado, F., Nogueira, N., Rato, L., Rijo, M.: Adaptive and non-adaptive model predictive control of an irrigation channel. Networks and Heterogeneous Media 4(2), 303–324 (2009)
12. Choy, S., Weyer, E.: Reconfiguration scheme to mitigate faults in automated irrigation channels. In: Proc. 44th IEEE Conf. Decision and Control, Sevilla, Spain, pp. 1875–1880 (2005)
13. Zhang, P.Z., Weyer, E.: A reference model approach to performance monitoring of control loops with applications to irrigation channels. Int. J. Adaptive Control Sig. Proc. 19(10), 797–818 (2005)
14. Liberson, D., Morse, A.S.: Basic problems in stability and design of switched systems. IEEE Control Systems 19(5), 59–70 (1999)

Estimation for Target Tracking Using a Control Theoretic Approach – Part I

Stephen C. Stubberud[1,*] and Arthur M. Teranishi[2]

[1] Oakridge Technology, Del Mar, CA, United States of America
scstubberud@ieee.org
[2] Asymmetric Associates, Long Beach, CA, United States of America
art_teranishi@cox.net

Abstract. In target tracking, the estimation problem is generated using the kinematic model of the target. The standard model is the straight-line motion model. Many variants to incorporate target maneuvers have been tried including interacting multiple motion models, adaptive Kalman filters, and neural extended Kalman filters. Each has performed well in a variety of situations. One problem with all of these approaches is that, without observability, the techniques often fail. In this paper, the first step in the development of a control-loop approach to the target-tracking problem is presented. In this effort, the use of a control law in conjunction with the estimation problem is examined. This approach is considered as the springboard for incorporating intelligence into the tracking problem without using ad hoc techniques that deviate from the underpinnings of the Kalman filter.

1 Introduction

One of the primary algorithms in a target tracking system is that of a state estimation routine [2]. When measurements are assigned to a specific target, they are incorporated into the target track via the estimation algorithm. The standard estimation algorithm is that of the Kalman filter or one of its many variants such as the extended Kalman filter (EKF) [5,9]. When the target is not maneuvering and fully observable measurements of the targets are provided, the Kalman filter provides a quality estimate of the target's kinematics, position and velocity. When the target maneuvers techniques such as the interactive multiple model (IMM) approach [3,10] and the neural extended Kalman filter [11, 12] have been employed to improve the performance of the tracking system. They improve the estimate and reduce delays in the measurements keeping up with the target.

When more complex issues occur, e.g., constraints and sensors that lack observability, often techniques are developed that are ad hoc in nature. These methods are developed in such a way that the authors claim that the technique is still an EKF even though it violates the basic tenets of the Kalman filter [14]. The estimation algorithm is dependent on the kinematic model of the target being tracked. The complexity of the model increases the implementation cost of the estimator. While a number of estimation algorithm have been developed such as the constrained estimator [13] and particle filters [6,8], they are still based on estimation theory.

* Corresponding author.

J. Świątek et al. (eds.), *Advances in Systems Science*,
Advances in Intelligent Systems and Computing 240,
DOI: 10.1007/978-3-319-01857-7_4,

Instead of modifying the estimator to solve the problem, in this paper, the use of a feedback control approach is used to improve the state estimate using the EKF algorithm. Instead of estimating the velocity vector of the target, a control approach is used to modify the velocity so as to better estimate the position of the target. This is a similar approach to that of satellite guidance [7]. The control approach is a better method to handle unobservability and constraints in that more complexities can be incorporated using control theory.

This paper is the first of three that investigates the use of control laws to improve the estimation of a target. In this paper, the estimation of the target's position for a straight-line and a maneuvering target is compared using a range-bearing sensor that uses the standard EKF approach to that of a control-based implementation of the estimator. This effort investigates if the technique can provide similar results to that of the EKF that uses a full state estimate when fully observable measurements are available. With the demonstration of the baseline system, the development of a system with unobservable measurements, i.e., an angle-only sensor, can be developed. The goal is to develop a system that can work well in standard tracking applications but then be modified to overcome the problems when unobservability occurs. The third step derives a more complex control law which can incorporate a number of rules and considerations.

In Section 2, the original EKF approach is derived and then modified for the use of the control law. Section 3 describes the exemplar cases. The results are presented and discussed in Section 4. The overview of the next steps in this research outlined in Section 5.

2 Extended Kalman Filter Designs

The standard EKF for tracking is based on a kinematic model of the target dynamics. As seen in Figure 1, there exist only internal feedback to the system. The prediction component is defined as

$$\mathbf{x}_{k+1|k} = \mathbf{F}\mathbf{x}_{k|k} \tag{1}$$

$$\mathbf{P}_{k+1|k} = \mathbf{F}\mathbf{P}_{k|k}\mathbf{F}^T + \mathbf{Q}_k \tag{2}$$

where the kinematic model, the state-coupling matrix **F**, is defined as

$$\mathbf{F} = \begin{bmatrix} 1 & dt & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & dt & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & dt \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

and the state vector **x** is

$$\mathbf{x}^T = \begin{bmatrix} x & \dot{x} & y & \dot{y} & z & \dot{z} \end{bmatrix} \tag{4}$$

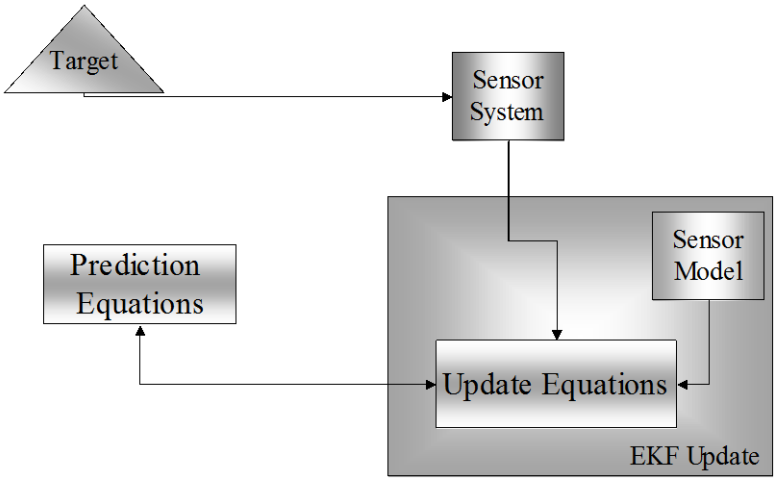

Fig. 1. The EKF is standard observer feedback loop with covariance information

This is a straight-line motion model. It is usually the best estimate of general target. Since the target is noncooperative in nature, any maneuver must be estimated through the process noise **Q.** There is no external input that can drive the system. The process noise for a tracker is often given as integrated white noise

$$\mathbf{Q} = q^2 \begin{bmatrix} \frac{dt^3}{3} & \frac{dt^2}{2} & 0 & 0 & 0 & 0 \\ \frac{dt^2}{2} & dt & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{dt^3}{3} & \frac{dt^2}{2}t & 0 & 0 \\ 0 & 0 & \frac{dt^2}{2} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{dt^3}{3} & \frac{dt^2}{2} \\ 0 & 0 & 0 & 0 & \frac{dt^2}{2} & dt \end{bmatrix} \tag{5}$$

which is defined in [1]. The variable q is defined based on an understanding of the target's capabilities. Large values are used for air targets while ground targets would use smaller values.

The update equations are given as

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}\right)^{-1} \tag{6}$$

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k \left(\mathbf{z}_k - \mathbf{h}\left(\mathbf{x}_{k|k-1} \right) \right) \tag{7}$$

$$\mathbf{P}_{k|k} = \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k \right) \mathbf{P}_{k|k} \tag{8}$$

The Jacobian **H** is a linearized version of the sensor dynamics **h**(•) in relation to the target states. The three main sensor types are those that provide azimuth/elevation (angle-only measurements)

$$\mathbf{h}\left(\mathbf{x}_{k|k-1} \right) = \begin{bmatrix} \alpha \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \arctan\left(\dfrac{\left(x_{tgt} - x_{platform} \right)}{\left(y_{tgt} - y_{platform} \right)} \right) \\ \arctan\left(\dfrac{\left(z_{tgt} - z_{platform} \right)}{\sqrt{\left(x_{tgt} - x_{platform} \right)^2 + \left(y_{tgt} - y_{platform} \right)^2}} \right) \end{bmatrix} \tag{9}$$

range/bearing/elevation

$$\mathbf{h}\left(\mathbf{x}_{k|k-1} \right) = \begin{bmatrix} \rho \\ \alpha \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \sqrt{\left(x_{tgt} - x_{platform} \right)^2 + \left(y_{tgt} - y_{platform} \right)^2 + \left(z_{tgt} - z_{platform} \right)^2} \\ \alpha \\ \varepsilon \end{bmatrix} \tag{10}$$

and range/bearing/elevation/range-rate

$$\mathbf{h}\left(\mathbf{x}_{k|k-1} \right) = \begin{bmatrix} \rho \\ \alpha \\ \varepsilon \\ \dot{\rho} \end{bmatrix} = \begin{bmatrix} \rho \\ \alpha \\ \varepsilon \\ \dfrac{x}{\rho}\dot{x} + \dfrac{y}{\rho}\dot{y} + \dfrac{z}{\rho}\dot{z} \end{bmatrix} \tag{11}$$

The measurement error covariance **R** is based on the sensor accuracy for the measurement **z**.

For the control law approach, the primary difference is in the prediction equations. The standard equations of the EKF with a controller become

$$\mathbf{x}_{k+1|k} = \mathbf{F}\mathbf{x}_{k|k} + \mathbf{g}\left(\mathbf{u}_k \right) \tag{12}$$

$$\mathbf{P}_{k+1|k} = \mathbf{F}\mathbf{P}_{k|k}\mathbf{F}^T + \mathbf{G}E\left[\mathbf{u}\mathbf{u}^T \right]\mathbf{G}^T + \mathbf{Q}_k \tag{13}$$

This incorporates an external input which as seen in Figure 2 allows us to incorporate a control law into the estimation effort. The new state vector becomes

$$\mathbf{x}^T = \begin{bmatrix} x & y & z \end{bmatrix} \tag{14}$$

with a kinematic equation

$$\mathbf{F} = \mathbf{I}_{3x3} \tag{15}$$

The control law becomes the velocity component of the estimation problem. If the range-rate is measurement it can be used to drive the position estimate.

3 Target Tracking Examples

Two basic target tracking problems are used to demonstrate the capabilities of the basic approach of this estimation routine.

3.1 Straight-Line Target Using Range-Bearing Measurements

The first target problem is a straight-line target with constant velocity. Figure 3 shows the target motion (dashed line) along with the sensor platform and its trajectory (solid dot line). A subset of the measurements is seen as solid lines with a diamonds on the end. The measurements are range-bearing measurements. The range error is assumed to be 1m, and the bearing error is assumed to 0.1 radians. The process noise factor, q, is set to 0.17 [1]. The time between measurements is 1.0 seconds.

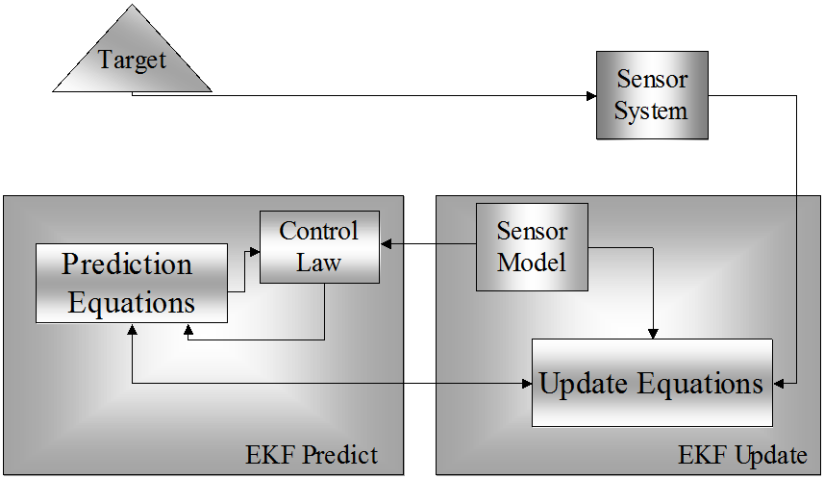

Fig. 2. With the control law, the EKF incorporates information in manners that can be much different than those of the standard estimator

The control law will use the variation in the previous updated estimate and the conversion of the measurement to the position state coordinate frame (the inverse function of state-to-measurement) and divide by the time difference between the last measurement and the current measurement:

$$\begin{bmatrix} \rho_{k+1} \sin \eta_{k+1} \\ \rho_{k+1} \cos \eta_{k+1} \end{bmatrix} \frac{1}{dt} - \mathbf{x}_{k|k} \tag{16}$$

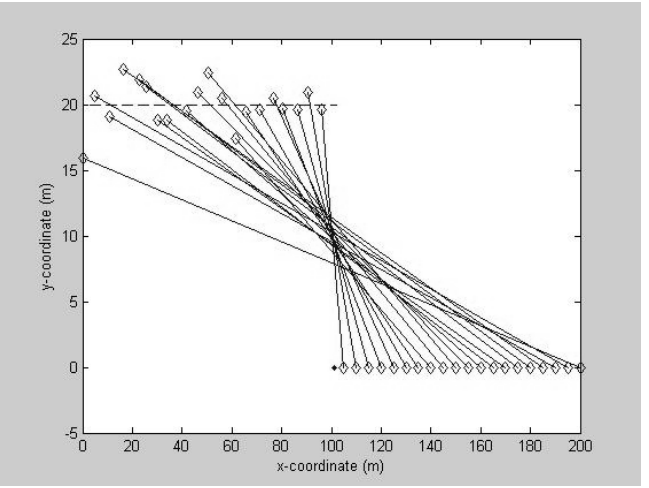

Fig. 3. The straight-line motion target. The sensor platform is lower right headed to the left with the target tracked north and west.

3.2 Maneuvering Target Using Range-Bearing-Range-Rate Measurements

The second example is that of target performing a straight-line target that is followed by a turn which the target again starts a straight-line leg as seen in Figure 4. Again, the target is denoted in the figure as the dash line. The platform is denoted by the solid-dot line. Again, the measurements are seen as solid lines with diamonds. The measurement will have a range-rate component as well with a measurement error of 0.1m/s. Otherwise, the sensor is the same. Based on implementation knowledge [4], the Jacobian for the range rate will have its position components zeroed out. The other errors and update are the same as mention in Section 3.1.

The control law uses the range rate to calculate a weight between the velocity created by the estimates and the velocity created by the estimated position and the position created by the next measurement

$$\begin{aligned} \mathbf{u}_1 &= \left(\begin{bmatrix} \rho_{k+1} \sin \eta_{k+1} \\ \rho_{k+1} \cos \eta_{k+1} \end{bmatrix} - \mathbf{x}_{k|k} \right) \frac{1}{dt} \\ \mathbf{u}_{21} &= \frac{\mathbf{x}_{k+1|k} - \mathbf{x}_{k|k-1}}{dt} \\ \mathbf{u} &= \left(w_1 \mathbf{u}_1 + w_2 \mathbf{u}_2 \right) \Big/ \left(w_1 + w_2 \right) \end{aligned} \tag{17}$$

where

$$w_1 = \left| \frac{\rho_{k+1} \sin\eta_{k+1}}{\left(\left(\rho_{k+1}\sin\eta_{k+1} - x_{1,ownship}\right) + \left(\rho_{k+1}\cos\eta_{k+1} - x_{2,ownship}\right)\right)^{1/2}} u_{1,1} + \left(\frac{\rho_{k+1}\cos\eta_{k+1}}{\left(\left(\rho_{k+1}\sin\eta_{k+1} - x_{1,ownship}\right) + \left(\rho_{k+1}\cos\eta_{k+1} - x_{2,ownship}\right)\right)^{1/2}} u_{1,2} \right) - \dot{\rho}_{k+1} \right| \tag{18}$$

$$w_2 = \left| \frac{x_{1,k+1|k}}{\left(\left(x_{1,k+1|k} - x_{1,ownship}\right) + \left(x_{2,k+1|k} - x_{2,ownship}\right)\right)^{1/2}} u_{2,1} + \left(\frac{x_{2,k+1|k}}{\left(\left(x_{1,k+1|k} - x_{1,ownship}\right) + \left(x_{2,k+1|k} - x_{2,ownship}\right)\right)^{1/2}} u_{2,2} \right) - \dot{\rho}_{k+1} \right| \tag{19}$$

4 Results

Both target tracking examples were processed through the standard EKF and the control version. A single run for each using the same noisy data was applied. The goal of this effort was to determine if a control-law-based approach could be made to provide similar results to that the standard EKF approach that uses a velocity component in the state vector. The results were compared against truth. In Figure 5, the absolute position-error was generated for both techniques applied to the nonmaneuvering target. The control approach error is denoted as the dashed line while the standard approach is shown as the solid line. The results are slightly offset but both appear to be well within the combined range and cross-range errors.

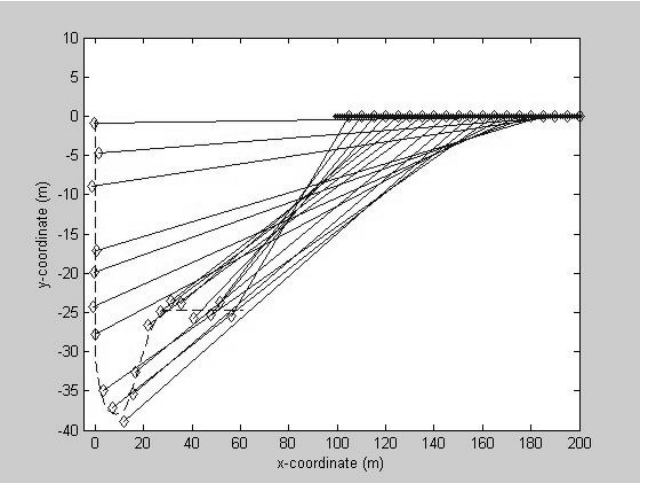

Fig. 4. The maneuver target example. The sensor platform is in the upper right with the target pulling a maneuver.

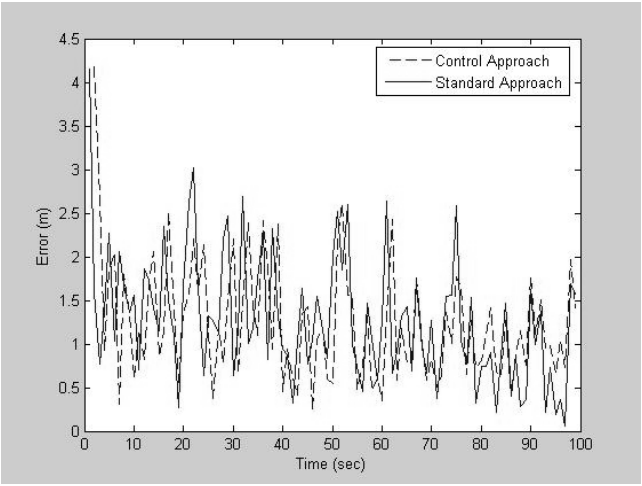

Fig. 5. Absolute position-error using both estimation techniques show similar results for the straight-line target tracking problem

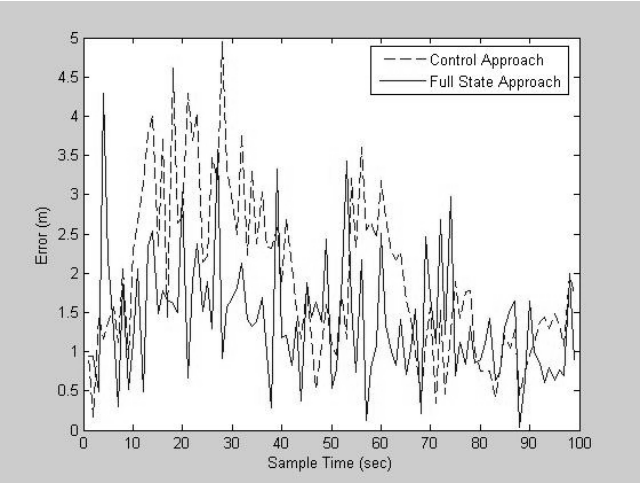

Fig. 6. Absolute position-error using both estimation techniques for the maneuvering target case shows that the control technique performs slightly worse until near the end of the scenario

In Figure 6, the absolute position-error of the tracking problem with the maneuvering target is shown. Here, the control-law approach performs slightly worst in this single case than the standard EKF approach. The control law approach still remains within the combined measurement errors (range and cross-range). Although this analysis has shown that the standard approach is better, it a single case and also indicate that the control approach will provide a comparable result in the cases where the tracker has fully observable measurements.

5 Conclusions and Future Directions

The first step in this research has shown that a control-law estimation technique can be used in a fully observable tracking problem. This allows the estimation routine to be general enough that it can be a generic tracking system. When the new algorithms for angle-only tracking and range-only tracking, the approach does not have to be used in an ad hoc manner where the technique is transitioned to when the issues with tracking arise. Analysis of this approach will be continued with Monte Carlo runs of these and different target trajectories.

In the next step of this research effort, a new control law will be developed for the angle-only tracking problem. This will be followed by a fuzzy-control algorithm that can switch between different tracking issues seamlessly.

References

1. Bar-Shalom, Y., Li, X.-R.: Estimation and Tracking: Principles, Techniques, and Software. Artech House Inc., Norwood (1993)
2. Blackman, S.: Multiple-Target Tracking with Radar Applications. Artech House, Norwood (1986)
3. Blackman, S., Popoli, R.: Design and Analysis of Modern Tracking Systems. Artech House, Norwood (1999)
4. Blackman, S.: Personal Correspondence (2013)
5. Brown, R.G., Hwang, P.Y.C.: Introduction to Random Signals and Applied Kalman Filters, 4th edn. Wiley, New York (2012)
6. Cappe, O., Godsill, S., Moulines, E.: An overview of existing methods and recent advances in sequential Monte Carlo. Proceedings of IEEE 95(5), 899 (2007)
7. Clements, R., Tavares, P., Lima, P.: Small Satellite Attitude Control Based a the Kalman Filter. In: Proceeding of the 2000 IEEE Symposium on Intelligent Control, Yokahama, Japan, pp. 79–84 (September 2000)
8. Doucet, A., Godsill, S., Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and Computing 10(3), 197–208 (2000)
9. Haykin, S.: Kalman Filters and Neural Networks. John Wiley & Sons (2004)
10. Kirubarajan, T., Bar-Shalom, Y.: Tracking Evasive Move-Stop-Move Targets With A GMTI Radar Using A VS-IMM Estimator. IEEE Transactions on Aerospace and Electronic Systems 39(3), 1098–1103 (2003)
11. Kramer, K.A., Stubberud, S.C.: Tracking of multiple target types with a single neural extended Kalman filter. International Journal of Intelligent Systems 25(5), 440–459 (2010)
12. Owen, M.W., Stubberud, A.R.: NEKF IMM Tracking Algorithm. In: Drummond, O. (ed.) Proceedings of SPIE: Signal and Data Processing of Small Targets 2003, San Diego, California, vol. 5204. TBD (August 2003)
13. Stubberud, A.R.: Constrained Estimate of the State of a Time-Variable System. In: Proceedings of the 14th International Conference on Systems Engineering, Coventry, UK, pp. 513–518 (September 2000)
14. Yang, C., Blasch, E.: Fusion of Tracks with Road Constraints. Journal of Advances of Information Fusion 3(1), 14–31 (2008)

LQ Optimal Control of Periodic Review Perishable Inventories with Transportation Losses

Piotr Lesniewski and Andrzej Bartoszewicz

Institute of Automatic Control, Technical University of Lodz, 90-924 Lodz,
18/22 Bohdana Stefanowskiego St., Poland
piotr.lesniewski2@gmail.com, andrzej.bartoszewicz@p.lodz.pl

Abstract. In this paper an LQ optimal warehouse management strategy is proposed. The strategy not only explicitly takes into account decay of commodities stored in the warehouse (perishable inventory) but it also accounts for transportation losses which take place on the way from supplier to the warehouse. The proposed strategy ensures full customers' demand satisfaction and prevents from exceeding the warehouse capacity. Moreover, it guarantees that the ordered quantities of goods are bounded and it helps achieve good trade-off between fast reaction of the system to time-varying demand and the big volume of the ordered goods. These favourable properties of the proposed strategy are formally stated as three theorems and proved in the paper.

Keywords: LQ optimal control, discrete time systems, inventory control, perishable inventory.

1 Introduction

The control theoretic approach to the issue of supply chain management has recently become an important research subject. An overview of the techniques used in the field and the obtained results can be found in [1-4]. The first application of the control theory methods to the management of logistic processes was reported in the early 1950s when Simon [5] applied servomechanism control algorithm to find an efficient strategy of goods replenishment in continuous time, single product inventory control systems. A few years later the discrete time servomechanism control algorithm for the purpose of efficient goods replenishment has been proposed [6]. Since that time numerous solutions have been presented, and therefore, further in this section we are able to mention only a few, arbitrarily selected examples of solutions proposed over the last decades. In [7] and [8] autoregressive moving average (ARMA) system structure has been applied in order to model uncertain demand. Then in [9] and [10] model predictive control of supply chain has been proposed and in [11] a robust controller for the continuous-time system with uncertain processing time and delay has been designed by minimising H_∞-norm. However, practical implementation of the strategy described in [11] requires application of numerical methods in order to obtain the control law parameters, which limits its analytical tractability.

In [12] lead-time delay for conventional, non-deteriorating inventories is explicitly taken into account and represented by additional state variables in the state space description. This approach results in the optimal controller designed by minimisation of

J. Świątek et al. (eds.), *Advances in Systems Science*,
Advances in Intelligent Systems and Computing 240,
DOI: 10.1007/978-3-319-01857-7_5,

quadratic performance index. An extension of the results presented in [12] to the case of perishable inventories is given in [13]. However, the analysis given in [13] does not take into account transportation losses (or in other words goods decay during the order procurement time). Therefore, in this paper we consider perishable inventories and we explicitly account for the ordered goods losses during the non-negligible lead time.

In this paper we consider a periodic-review inventory system with perishable goods replenished from a single supply source. Contrary to the previously published results we consider not only losses which take place when the commodity is stored in the warehouse, but also those which happen during the supply process, i.e. the losses on the way from the supplier to the warehouse. We propose a discrete time representation of the supply chain dynamics and we apply minimization of the quadratic performance index to design the controller for the considered system. This index takes into account not only the stock level, but also the amount of goods en route to the distribution center, which has not been considered in earlier works. The controller is determined analytically in a closed form, which allows us to state and formally prove important properties of proposed inventory policy. First, we prove that the designed management policy always generates strictly positive and upper bounded order quantities, which is an important issue from the practical point of view. Next, we define the warehouse capacity which provides enough space for all incoming shipments, and finally we derive a condition which must be satisfied in order to guarantee 100% service level, i.e. full satisfaction of imposed demand.

2 Inventory Replenishment System Model

In this paper we consider a periodic review inventory replenishment system with an unknown, time-varying demand $d(kT)$ and transportation losses. The inventory is replenished from a distant supply source with a lead time L. The lead time is a multiple of the review period T, i.e. $L = mT$, where m is a positive integer. The model of the inventory replenishment system considered in this paper is illustrated in Figure 1. The amount of goods ordered at time kT (where $k = 0, 1, 2, ...$) is denoted by $u(kT)$. The orders are determined using the current stock level $y(kT)$, the demand stock level y_d and the order history. Since we explicitly take into account transportation losses, only αu of goods (where $0 < \alpha \leq 1$) reach the warehouse. Furthermore, as we consider perishable commodities, during each review period a fraction σ ($0 \leq \sigma < 1$) of the stock deteriorates. The demand is modeled as an *a priori* unknown, nonnegative function of time $d(kT)$, its upper bound is denoted by d_{max}. If there are enough items in stock, the demand is fully covered, otherwise, only a part of the demand is satisfied. Therefore, we introduce an additional function $h(kT)$ which represents the amount of goods actually sold at review period k. Thus

$$0 \leq h(kT) \leq d(kT) \leq d_{max}. \tag{1}$$

The current stock level can be presented in the following form

$$y[(k+1)T] = py(kT) + \alpha u[(k-m)T] - h(kT). \tag{2}$$

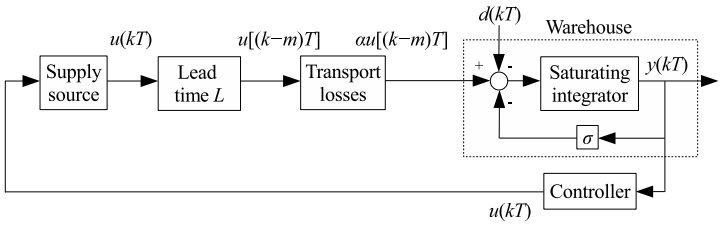

Fig. 1. Inventory system model

where $p = 1 - \sigma$ is the fraction of the stock remaining in the warehouse. Of course, since $0 \leq \sigma < 1$ we have $0 < p \leq 1$. We assume that the warehouse is initially empty, $y(0) = 0$, and the first order is placed at $k = 0$. The first order arrives at the warehouse at mT, and $y(kT) = 0$ for $k \leq m$. We assume, that the goods (apart from the fraction $1 - \alpha$ which is broken during transport) reach the stock new, and deteriorate while kept in it. The stock level can be rewritten in the following form

$$y(kT) = \alpha \sum_{j=0}^{k-m-1} p^{k-m-1-j} u(jT) - \sum_{j=0}^{k-1} p^{k-1-j} h(jT). \tag{3}$$

In order to apply a control theoretic approach to this problem it is useful to represent the model in the state space. We select the first variable as the stock level, $x_1(kT) = y(kT)$. The remaining state variables represent the delayed values of the control signal, i.e. $x_j(kT) = u(k - n + j - 1)$ for $j = 2, \ldots, n$, where n is the system order. We can now describe the system in the state space as

$$\begin{aligned} \boldsymbol{x}[(k+1)T] &= \boldsymbol{A}\boldsymbol{x}(kT) + \boldsymbol{b}u(kT) + \boldsymbol{o}h(kT) \\ y(kT) &= \boldsymbol{q}^T \boldsymbol{x}(kT), \end{aligned} \tag{4}$$

where $\boldsymbol{A}$ is a n x n state matrix and $\boldsymbol{b}$, $\boldsymbol{o}$, $\boldsymbol{q}$ are n x 1 vectors

$$\boldsymbol{A} = \begin{bmatrix} p & \alpha & 0 & & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & & 0 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{o} = \begin{bmatrix} -1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{q} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}. \tag{5}$$

Since the goods perish at rate σ when kept in the warehouse, in order to keep the stock at the demand level y_d it is necessary to constantly refill it at rate σy_d. Taking into account transport losses the desired system state is therefore given by

$$\boldsymbol{x_d} = y_d \begin{bmatrix} 1 & \sigma/\alpha & \ldots & \sigma/\alpha \end{bmatrix}^T. \tag{6}$$

3 Proposed Supply Management Strategy

In this section we will develop a LQ optimal controller for the considered inventory system. Its important properties will then be formulated and proved.

In optimization problems we often consider a quadratic quality criterion that involves the control signal and the state error vector $\boldsymbol{e} = \boldsymbol{x_d} - \boldsymbol{x}$. Also in this paper we seek for a control law that minimizes the following cost functional

$$J(u) = \sum_{k=0}^{\infty} \left[u^2(kT) + \boldsymbol{e}^T(kT) \boldsymbol{W} \boldsymbol{e}(kT) \right]. \tag{7}$$

We choose $\boldsymbol{W} = \mathrm{diag}(w_1, w_2, \ldots, w_2)$ where w_1 and w_2 are positive coefficients adjusting the influence of the stock level error and the amount of goods in transit respectively. This quality criterion is more general than the one presented in [13], which simply ignored the amount of goods currently held in transport. According to [14] the optimal control $u_{opt}(kT)$ that minimizes the cost functional (7) can be presented as

$$u_{opt}(kT) = r - \boldsymbol{g}\boldsymbol{x}(kT), \tag{8}$$

where

$$\boldsymbol{g} = \boldsymbol{b}^* \boldsymbol{K} \left(\boldsymbol{I_n} + \boldsymbol{b}\boldsymbol{b}^* \boldsymbol{K} \right)^{-1} \boldsymbol{A}, \tag{9}$$

and r is a constant term. Operator $(.)^*$ denotes the complex conjugate matrix transpose, semipositive matrix $\boldsymbol{K}$ satisfies $\boldsymbol{K}^* = \boldsymbol{K}$ and is determined by the following Ricatti equation

$$\boldsymbol{K} = \boldsymbol{A}^* \boldsymbol{K} \left(\boldsymbol{I_n} + \boldsymbol{b}\boldsymbol{b}^* \boldsymbol{K} \right)^{-1} \boldsymbol{A} + \boldsymbol{W}. \tag{10}$$

Because all elements of $\boldsymbol{A}$, $\boldsymbol{b}$ and $\boldsymbol{q}$ are real numbers, the complex conjugate matrix transpose $(.)^*$ is equivalent to the matrix transpose $(.)^T$. This means, that all elements of matrix $\boldsymbol{K}$ are also real numbers. Therefore, condition $\boldsymbol{K}^* = \boldsymbol{K}$ implies that $\boldsymbol{K}$ is symmetric. Because of this, in order to make notation as concise as possible, we will represent the elements of $\boldsymbol{K}$ below the main diagonal by '$*$'.

As the system order n depends on the transport delay, in order to draw general conclusions we need to solve (10) analytically for an arbitrary system order. The approach proposed here is similar to the one used in [13] and involves iterative substitution of $\boldsymbol{K}$ into the right hand side of equation (10) and comparing with its left hand side, so that at each iteration the number of independent elements of $\boldsymbol{K}$ is reduced.

We begin by substituting the most general form of $\boldsymbol{K}$ into (10) and obtain

$$\boldsymbol{K}_1 = \begin{bmatrix} k_{11} & \frac{\alpha}{p}(k_{11} - w_1) & k_{13} & k_{1n} \\ * & \frac{\alpha^2}{p^2}(k_{11} - w_1) + w_2 & k_{23} \cdots & k_{2n} \\ * & * & k_{33} & k_{3n} \\ & \vdots & \ddots & \vdots \\ * & * & * & k_{nn} \end{bmatrix}. \tag{11}$$

The next step involves substituting (11) into (10) and comparing the left and right hand sides. We then repeat this procedure, until all elements of $\boldsymbol{K}$ are expressed as functions of k_{11}, system parameters p and α, system order n and the weighting factors w_1 and w_2

$$\boldsymbol{K} = \begin{bmatrix} k_{11} & \frac{\alpha}{p}(k_{11}-w_1) & \cdots & \frac{\alpha}{p^{n-1}}\left(k_{11} - w_1 \sum_{i=0}^{n-2} p^{2i}\right) \\ * & \frac{\alpha^2}{p^2}(k_{11}-w_1)+w_2 & \cdots & \frac{\alpha^2}{p^n}\left(k_{11} - w_1 \sum_{i=0}^{n-2} p^{2i}\right) \\ & \vdots & \ddots & \vdots \\ * & * & & \frac{\alpha^2}{p^{2n-2}}\left(k_{11} - w_1 \sum_{i=0}^{n-2} p^{2i}\right) + (n-1)w_2 \end{bmatrix}. \tag{12}$$

Now we can determine k_{11} by substituting (12) into (10) and comparing the first elements of the obtained matrices. This results in

$$k_{11}\left\{w_1\alpha^2\left(2 - p^{2n-2} + 2p^2\frac{1-p^{2n-2}}{1-p^2}\right) - p^{2n-2}(1-p^2)[1+(n-1)w_2]\right\} + \\ - \frac{w_1^2\alpha^2(1-p^{2n}-p^{2n-2}+p^{4n-2})}{(1-p^2)^2} + p^{2n-2}w_1[1+(n-1)w_2] - \alpha^2 k_{11}^2 = 0. \tag{13}$$

Equation (13) has two roots

$$k_{11}^{\pm} = \frac{w_1\alpha^2\left(\frac{2-p^{2n}-p^{2n-2}}{1-p^2}\right) - p^{2n-2}(1-p^2)[1+(n-1)w_2] \pm p^{2n-2}\sqrt{\Delta}}{2\alpha^2}, \tag{14}$$

where

$$\Delta = w_1^2\alpha^4 + 2\alpha^2 w_1[1+(n-1)w_2](p^2+1) + (1-p^2)^2[1+(n-1)w_2]^2. \tag{15}$$

Only k_{11}^{+} guarantees that $\boldsymbol{K}$ is semipositive definite. Having found $\boldsymbol{K}$, using k_{11}^{+} with (9) we derive vector $\boldsymbol{g}$

$$\boldsymbol{g} = \gamma\left[1/\alpha \; 1/p \; 1/p^2 \; \ldots \; 1/p^{n-1}\right], \tag{16}$$

where

$$\gamma = p^n\left\{1 - \frac{2[1+(n-1)w_2]}{w_1\alpha^2 + [1+(n-1)w_2](1+p^2) + \sqrt{\Delta}}\right\}. \tag{17}$$

We can now observe, that in order for the state to reach $\boldsymbol{x_d}$ defined by (6)

$$r = y_d\left[1 - p + \gamma p^{-(n-1)}\right]/\alpha. \tag{18}$$

As all state variables except x_1 are the delayed values of the control signal, we conclude, that the control signal of the LQ optimal controller is given by

$$u_{opt}(kT) = r - \frac{\gamma y(kT)}{\alpha} - \gamma p^{-n}\sum_{i=k-m}^{k-1} p^{k-i}u(iT). \tag{19}$$

This completes the design of the LQ optimal controller. Next we will present and prove important properties of the proposed control strategy.

3.1 Stability Analysis

The design method applied in this work ensures stability of the closed loop system. In order to verify this property let us notice that a discrete time system is asymptotically stable if all the roots of its closed loop state matrix $\boldsymbol{A_c}$ lie inside the unit circle on the z-plane. In our case $\boldsymbol{A_c} = \boldsymbol{A} - \boldsymbol{bg}$, and the characteristic polynomial has the following form

$$det(z\boldsymbol{I_n} - \boldsymbol{A_c}) = z^{n-1}[z - p(1 - \gamma p^{-n})]. \tag{20}$$

All roots of (20) are located inside the unit circle if $-1 < p(1 - \gamma p^{-n}) < 1$. It can be seen from (17) that $0 < \gamma < p^n$. Since $0 < p \leq 1$, we conclude, that all but one roots of (20) lie in the origin of the z-plane, and one root lies between 0 and p, depending on the value of γ. Therefore, the closed loop system is stable and no oscillations appear at the output.

Remark 1. Weighting factors w_1 and w_2 can be tuned to meet specific requirements. When $w_1 \to 0$ the value of the control signal dominates the quality criterion, and gain γ drops to zero. When $w_1 \to \infty$ for any finite w_2 the output error is to be reduced to zero as quickly as possible, no matter the value of the control signal. The controller then becomes a dead-beat scheme, its gain γ approaches p^n. Errors of states $x_2, x_3, \ldots, x_n$ represent the difference between the current replenishment orders and the steady-state order $(1 - p)y_d/\alpha$. Therefore, increasing w_2 leads to decrease of γ and vice versa.

3.2 Properties of the Proposed Controller

In this section the properties of the control strategy proposed in this paper will be stated in three theorems. In the first one we will show that generated order quantities are always nonnegative and upper bounded. The second theorem will specify the warehouse capacity needed to always accommodate the incoming shipments. The last theorem will show how to select the demand stock value in order to ensure full consumer demand satisfaction.

Theorem 1. *The order quantities generated by the control strategy* (19) *are always bounded, and satisfy the following inequality*

$$u_m \leq u(kT) \leq \max(r, u_M), \tag{21}$$

where

$$u_m = \frac{r(1-p)}{1 + p(\gamma p^{-n} - 1)}, \quad u_M = \frac{r(1-p) + \gamma d_{max}/\alpha}{1 + p(\gamma p^{-n} - 1)}. \tag{22}$$

Proof. It follows from (19) *that* $u(0) = r$. *Therefore,* (21) *is satisfied for* $k = 0$. *Substituting* (3) *into* (19) *we obtain*

$$u(kT) = r - \frac{\gamma}{\alpha}\left[\alpha \sum_{j=0}^{k-1} p^{k-m-1-j} u(jT) - \sum_{j=0}^{k-1} p^{k-1-j} h(jT)\right]. \tag{23}$$

Now we assume, that (21) *holds for all integers up to some* $l \geq 0$*. We will show that this implies that* (21) *is also true for* $l+1$*. We can rewrite* (23) *for* $k = l+1$ *as follows*

$$u[(l+1)T] = r - \gamma p^{-m} u(lT) + \frac{\gamma}{\alpha} h(lT) - \frac{\gamma}{\alpha}\left[\alpha \sum_{j=0}^{l-1} p^{l-m-j} u(jT) - \sum_{j=0}^{l} p^{l-j} h(jT)\right]. \tag{24}$$

The last term in the above equation is equal to $p[r - u(l)]$*. Consequently*

$$u[(l+1)T] = r(1-p) + p(1-\gamma p^{-n}) u(lT) + \gamma h(lT)/\alpha. \tag{25}$$

Because $h(kT)$ *is always nonnegative we can obtain from* (25) *the minimum value of the control signal*

$$u(lT) \geq r(1-p) \Big/ [1 + p(\gamma p^{-n} - 1)], \tag{26}$$

which shows that the first inequality in (21) *actually holds. Since* $h(kT) \leq d_{max}$ *for any* $k \geq 0$ *we can calculate the maximum value of the control signal from* (25) *as*

$$u[(l+1)T] \leq r(1-p) + p(1-\gamma p^{-n}) u(lT) + \gamma d_{max}/\alpha. \tag{27}$$

First we will consider the case when $r \geq u_M$*. From* (22) *we obtain*

$$\gamma d_{max}/\alpha \leq \gamma p^{-m} r. \tag{28}$$

Using this relation with (27) *we arrive at*

$$u[(l+1)T] \leq r(1-p) + p(1-\gamma p^{-n}) r + \frac{\gamma}{\alpha} d_{max} = r - \gamma p^{-m} r + \frac{\gamma}{\alpha} d_{max} \leq r. \tag{29}$$

For the second case, when $r < u_M$ *from* (25) *we get*

$$u[(l+1)T] \leq r(1-p) + p(1-\gamma p^{-n}) \frac{r(1-p) + \gamma d_{max}/\alpha}{1 + p(\gamma p^{-n} - 1)} + \gamma d_{max}/\alpha = u_M. \tag{30}$$

Taking into account relations (26)*,* (29) *and* (30) *and using the principle of mathematical induction we conclude, that* (21) *indeed holds for any* $k \geq 0$*. This ends the proof.*

In real inventory systems it is necessary to ensure a finite warehouse size, that will always accommodate the incoming shipments. The next theorem shows, that application of our strategy ensures that the stock level will never exceed a precisely determined, *a priori* known value.

Theorem 2. *If the proposed control strategy is applied, then the on-hand stock will never exceed its demand value, i.e. for any* $k \geq 0$

$$y(kT) \leq y_d. \tag{31}$$

Proof. The warehouse is empty for any $k \leq m = n-1$*. Therefore, we only need to show that* (31) *holds for* $k \geq n$*. We begin by assuming, that* (31) *holds for some integer*

$l \geq n$. We will demonstrate, that this assumption implies $y[(l+1)T] \leq y_d$. Applying (23) *to* (2) *we get*

$$y[(l+1)T] = py(lT) + \alpha r - \gamma p^{-n} \sum_{j=l-m}^{l-1} p^{l-j} h(jT) - h(lT) + \\ - \gamma p^{-m} \left[\alpha \sum_{j=0}^{l-m-1} p^{l-m-1-j} u(jT) - \sum_{j=0}^{l-1} p^{l-1-j} h(jT) \right]. \quad (32)$$

We observe from (3)*, that the terms in the square brackets are equal to the on-hand stock level in period l. Consequently*

$$y[(l+1)T] = \alpha r + py(lT)(1 - \gamma p^{-n}) - \gamma p^{-n} \sum_{j=l-m}^{l-1} p^{l-j} h(jT) - h(lT). \quad (33)$$

We assumed, that $y(lT) \leq y_d$, the amount of sold goods $h(kT)$ is always nonnegative, and r is given by (18)*. Thus, from* (33) *we can obtain*

$$y[(l+1)T] \leq y_d(1 - p + \gamma p^{-(n-1)}) + y_d(p - \gamma p^{-(n-1)}) = y_d. \quad (34)$$

We conclude, using the principle of mathematical induction, that (31) *is true for $k \geq n$. As stated before, $y(kT) = 0$ for $k < n$. Therefore* (31) *holds for any $k \geq 0$.*

It follows from Theorem 2, that if we assign a storage capacity equal to y_d at the distribution center, then all incoming shipments will be accommodated, and thus the high cost of emergency storage is eliminated. A successful inventory policy is also required to ensure high demand satisfaction. The last theorem will provide a formula for the smallest possible y_d that always ensures 100% consumer demand satisfaction.

Theorem 3. *With the application of the proposed control law, if the reference stock level satisfies inequality*

$$y_d > d_{max} \frac{1 + \gamma \sum_{j=-m}^{1} p^j}{1 - p + \gamma p^{-(n-1)}}, \quad (35)$$

then for any $k \geq m+1$ the stock level is strictly positive.

Proof. We will show that if (35) *holds, then for any $l \geq m+1$ condition $y[(l-1)T] > 0$ implies $y(lT) > 0$. Using* (1) *with* (33) *for $l \geq m+1$ we get*

$$y[(l+1)T] \geq y_d[1 - p + \gamma p^{-(n-1)}] - d_{max} \left(1 + \gamma \sum_{j=-m}^{-1} p^j \right) > 0. \quad (36)$$

The stock level $y(kT) = 0$ for $k \leq m$. Therefore, we can obtain the stock size for $k = m+1$ from (2) *as*

$$y[(m+1)T] = \alpha r - h(mT) > 0. \quad (37)$$

Again using the principle of mathematical induction with (36) *and* (37) *we conclude, that indeed if y_d satisfies* (35)*, then the stock level is positive for any $k \geq m+1$.*

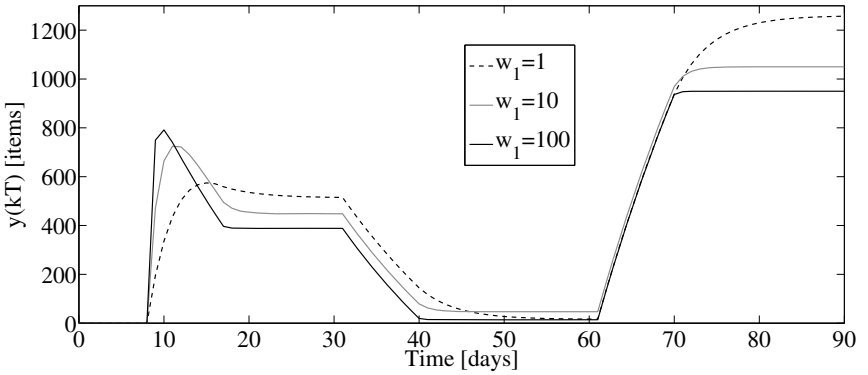

Fig. 2. Stock level for different values of w_1

We notice from (2), that a positive stock level in period k implies full satisfaction of demand in period $k-1$. Therefore, the above theorem demonstrates full satisfaction of demand $d(kT)$ for any $k \geq m$.

4 Simulation Results

In order to present the properties of the proposed control strategy, computer simulations are performed. The review period T is selected as 1 day. The lead time L is assumed to be 8 days. From this follows $m = 8$ and $n = 9$. Parameter $d_{max} = 120$ items. The actual demand is $d(kT) = 72$ for $k \in [0, 30)$, $d(kT) = 120$ for $k \in [30, 60)$ and $d(kT) = 0$ for $k \in [60, 90]$. Sudden changes of large amplitude occur in the function d, which reflects the most difficult conditions in the system. It is assumed, that 5% of the goods are broken during transport, which corresponds to $\alpha = 0.95$. The inventory decay factor $\sigma = 0.04$, which implies $p = 0.96$.

We select $w_2 = 2$ and perform three simulations, each one with a different value of w_1. The simulation parameters – gain γ obtained from (17), minimum stock demand level y'_d calculated from condition (35), and the demand stock level actually used in the simulation y_d are shown in Table 1. The results of the simulations are shown in figures 2 and 3. The value of control signal at the beginning of the transmission process is shown in Figure 4.

It can be seen from the figures, that the replenishment orders calculated by the proposed control law are always lower and upper bounded as stated in Theorem 1. Furthermore, the amount of goods in the warehouse never exceeds its demand value, and never

Table 1. Parameters of the LQ optimal controller

w_1	γ	y'_d	y_d
100	0.595	935	950
10	0.344	1031	1050
1	0.124	1244	1260

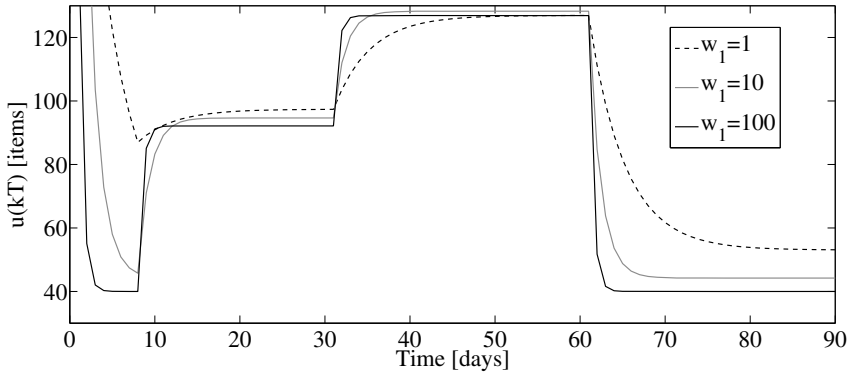

Fig. 3. Replenishment orders for different values of w_1

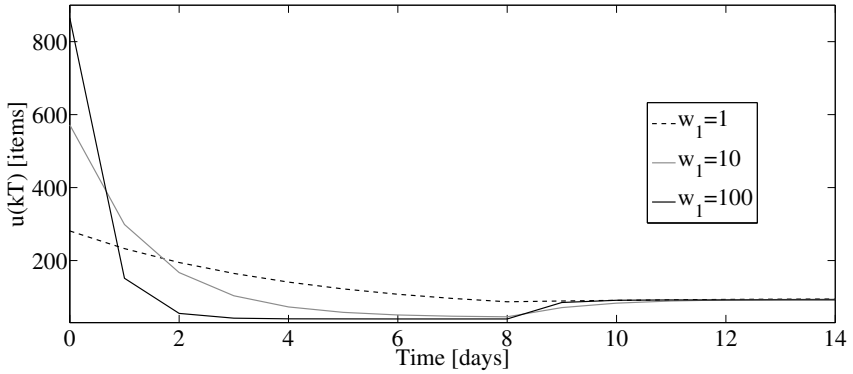

Fig. 4. Replenishment orders at the beginning of the control process

decreases to zero for $k \geq m + 1$. This means, that the incoming shipments are always accommodated in the distribution center, and that consumer demand is fully satisfied.

By selecting appropriate values of w_1 and w_2 we change the value of γ and can adapt the algorithm to particular needs. Larger values of γ result in faster tracking of consumer demand. This, in turn, allows allocating a smaller warehouse capacity, while still ensuring full consumer demand satisfaction. On the other hand, small γ leads to smaller replenishment orders at the beginning of the control process. It also makes the changes in replenishment orders smoother, which makes them easier to follow for the supplier.

5 Conclusions

In this paper an optimal periodic review supply chain management strategy has been proposed. The strategy takes into account perishable inventories with transportation losses, i.e. not only it explicitly concerns goods decay in the warehouse, but it also accounts for the losses which take place during the delivery process. The proposed strategy ensures full demand satisfaction, eliminates the risk of warehouse overflow

and always generates non-negative and bounded orders. The design procedure applied in this paper is based on minimization of a quadratic cost functional (which is more general than the similar ones proposed earlier [13]), and solving the resulting matrix Ricatti equation.

Acknowledgments. This work has been performed in the framework of a project "Optimal sliding mode control of time delay systems" financed by the National Science Centre of Poland decision number DEC 2011/01/B/ST7/02582. Kind support provided by the Foundation for Polish Science under Mistrz grant is also acknowledged.

References

1. Sarimveis, H., Patrinos, P., Tarantilis, C.D., Kiranoudis, C.T.: Dynamic modeling and control of supply chain systems: a review. Computers and Operations Research 35(11), 3530–3561 (2008)
2. Karaesmen, I., Scheller-Wolf, A., Deniz, B.: Managing perishable and aging inventories: review and future research directions. In: Kempf, K., Keskinocak, P., Uzsoy, R. (eds.) Handbook of Production Planning. Kluwer, Dordrecht (2008)
3. Boccadoro, M., Martinelli, F., Valigi, P.: Supply chain management by H-infinity control. IEEE Trans. on Automation Science and Engineering 5(4), 703–707 (2008)
4. Hoberg, K., Bradley, J.R., Thonemann, U.W.: Analyzing the effect of the inventory policy on order and inventory variability with linear control theory. European J. of Operations Research 176(3), 1620–1642 (2007)
5. Simon, H.A.: On the application of servomechanism theory in the study of production control. Econometrica 20(2), 247–268 (1952)
6. Vassian, H.J.: Application of discrete variable servo theory to inventory control. Arthur D. Little, Inc., Cambridge (1954)
7. Gaalman, G., Disney, S.M.: State space investigation of the bullwhip problem with ARMA(1,1) demand processes. International J. of Production Economics 104(2), 327–339 (2006)
8. Gaalman, G.: Bullwhip reduction for ARMA demand: the proportional order-up-to policy versus the full-state-feedback policy. Automatica 42(8), 1283–1290 (2006)
9. Aggelogiannaki, E., Doganis, P., Sarimveis, H.: An adaptive model predictive control configuration for production-inventory systems. International J. of Production Economics 114(1), 165–178 (2008)
10. Li, X., Marlin, T.E.: Robust supply chain performance via Model Predictive Control. Computers & Chemical Engineering 33(12), 2134–2143 (2009)
11. Boukas, E.K., Shi, P., Agarwal, R.K.: An application of robust technique to manufacturing systems with uncertain processing time. Optimal Control Applications and Methods 21(6), 257–268 (2000)
12. Ignaciuk, P., Bartoszewicz, A.: Linear-quadratic optimal control strategy for periodic-review inventory systems. Automatica 46(12), 1982–1993 (2010)
13. Ignaciuk, P., Bartoszewicz, A.: Linear-quadratic optimal control of periodic-review perishable inventory systems. IEEE Trans. on Control Systems Technology 20(5), 1400–1407 (2012)
14. Kwakernaak, H., Sivan, R.: Linear Optimal Control Systems. Wiley-Interscience, New York (1972)

A Dynamic Vehicular Traffic Control Using Ant Colony and Traffic Light Optimization

Mohammad Reza Jabbarpour Sattari, Hossein Malakooti, Ali Jalooli, and Rafidah Md Noor

Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

Abstract. Vehicle traffic congestion problem in urban areas due to increased number of vehicles has received increased attention from industries and universities researchers. This problem not also affects the human life in economic matters such as time and fuel consumption, but also affects it in health issues by increasing CO2 and greenhouse gases emissions. In this paper, a novel cellular ant-based algorithm combined with intelligent traffic lights based on streets traffic load condition has been proposed. In the proposed method road network will be divided into different cells and each vehicle will guide through the less traffic path to its destination using Ant Colony Optimization (ACO) in each cell. Moreover, a new method for traffic lights optimization is proposed in order to mitigate the traffic congestion at intersections. Two different scenarios have been performed through NS2 in order to evaluate our traffic lights optimization method. Based on obtained results, vehicles average speed, their waiting time and number of stopped vehicles at intersections are improved using our method instead of using usual traffic lights.

1 Introduction

Over the last decade, vehicle population has been increased sharply in the world. This large number of vehicles leads to a heavy traffic congestion and consequently,lots of accidents.According to RACQ Congested Roads report [1], fuel consumption, CO2 and greenhouse gases emissions, long travel time and accidents are both direct and indirect results of vehicle traffic congestion and rough (vs. smooth) driving pattern.

Accordingly, there should be a way to alleviate the vehicle congestion problem. Building new high capacity streets and highways can mitigate some of the aforementioned problems. Nevertheless, this solution is very costly,time consuming and in most of the cases, it is not possible because of the space limitations.On the other hand, optimal usage of the existent roads and streets capacity can lessen the congestion problem in large cities at the lower cost.However, this solution needs accurate information about current status of roads and streets which is a challenging task due to quick changes in vehicular networks and environments. Providing alternative paths with shortest time duration instead of shortest path distances can be useful because of lower fuel consumption and

J. Świątek et al. (eds.), *Advances in Systems Science*,
Advances in Intelligent Systems and Computing 240,
DOI: 10.1007/978-3-319-01857-7_6,

traffic congestion. These approaches are called Dynamic Traffic Routing System (DTRS). Various DTRSs are proposed in [2–4], but among them, using Multi Agent System (MAS) is reported as a promising and one of the best approaches for dynamic problems [5]. Particularly, ant agents have proven to be superior to other agents in [6–8].

In ant-based algorithms, inspired from real ants behavior, artificial ants (agents) find the shortest path from source to destination based on probabilistic search in the problem space. Dividing the routing space into several smaller spaces (cells) can lead to a better routing result because of dynamic nature of the vehicle's congestion. In addition to vehicle routing and traffic control, intersections can affect the traffic congestion and smooth driving pattern. This is because of traffic lights existence in the intersections. In addition, according to [9], up to 90% of the utilized traffic lights operate based on fixed assignments of green splits and cycle duration which leads to inessential stops of vehicles. Hence, optimizing the traffic lights can eliminate waste of time and money.

Therefore, we addressed some of aforementioned drawbacks by proposing a cellular ant-based algorithm applied to dynamic traffic routing , using optimized traffic lights for traffic congestion problem in vehicular environment. The rest of this paper in organized as follows: Section 2 discusses about related works in two different sections, dynamic vehicle routing using ACO and traffic lights optimization. Proposed methods for vehicle routing and traffic lights optimization are explained in Section 3. Obtained results are discussed and justified in Section 4. Section 5 concludes the paper.

2 Related Work

In this section most of the related approaches to our topic will be discussed. Since our approach has two parts, this section is divided into two subsections; dynamic traffic routing using ACO and traffic lights optimization. To the best of our knowledge, there is no approach which utilizes both of these approaches at the same time to reduce vehicle traffic congestion and our approach uses these two methods simultaneously for first time.

2.1 Traffic Light Optimization (TLO)

During last four decades, Urban Traffic Control (UTC) based on traffic light optimization has been attracted researchers and industries attention. Complex mathematical formulas and models are used in most of the existing UTC approaches in order to traffic lights optimization. SCATS (Sydney Coordinated Adaptive Traffic System) [10] and SCOOT (Split, Cycle and Offset Optimization Technique) [11] are the most well-known examples of this kind of UTC systems. adaptive traffic light approach based on wireless sensors is proposed in [12–14]. As compared to UTC system, by using these approaches more information such as vehicles direction, speed and location can be used for getting accurate decisions for TLO. Therefore, UTC systems problem which comes from

the fixed location of the detectors is solved in adaptive traffic light algorithms. In-vehicle Virtual Traffic Light (VTL) protocol is designed in [15] in order to traffic flow optimization at intersection without using road side infrastructure such as RSUs and traffic lights.

2.2 Dynamic Traffic Routing (DTR) using Ant Colony Optimization

Over dynamically changing networks, finding best routes can also be fulfilled through using Swarm Intelligence (SI) based methods. One of the most advantageous SI methods for exploring optimal solutions at low computational cost is ant routing algorithm. AntNet is a routing algorithm which is inspired by the natural ants behavior and operates based on distributed agents [8]. AntNet has been proved to be an adoptable algorithm to the changes in traffic flows and have better performance than other shortest path algorithms [7].Using ant colony algorithm in combination with network clustering autonomous system has been proved to be effective in finding best routing solutions by Kassabalidis et al. in [16]. Cooperation among neighboring nodes can be increased using a new type of helping ants which are introduced in [17]. Consequently, AntNet algorithms convergence time will be reduced as well. A new version of AntNet algorithm which improves the average delay and the throughput is introduced by Tekiner et al. in [18]. Moreover, the ant/packet ratio is used in their algorithm to constrain the number of using ants.

In road traffic routing, the significant role of Dynamic traffic routing algorithms to prevent facing congestion offer better routes to cars is noticeable. For car navigation in a city, a DTR which utilizes the Ant Based Control algorithm (ABC algorithm) is introduced in [2]. However, it is proved that this algorithm is more appropriate in small networks of city streets rather than big ones due to its scalability problems. An adjustment to the AntNet and the Ant Based Control (ABC) to direct drivers to the best routs by the aid of historically-based traffic information has been offered in [3].Another version of the AntNet algorithm by the help of which travel time can be improved over a congested network is presented in [19] and [20]. This improvement can be achieved through diverting traffic from congested routs.In hierarchical routing system (HRS), which is proposed in [4], roads are assigned to different hierarchy levels and consequently, a traffic network is split into several smaller networks or sectors. A routing table is for leading the cars to better routes is located at the networks intersections (at sector level and locally). For dynamic routing, an ant-based algorithm is utilized. The high adaptability of this approach in complex networks is noticeable.

3 Proposed Model

3.1 VANET based Traffic Light Optimization

In this section, we propose a new vehicle-to-traffic light counter based model which is used for traffic lights optimization as well as finding optimal path for

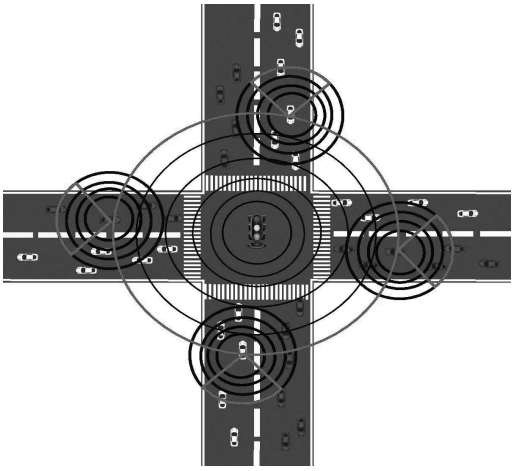

Fig. 1. Traffic light optimization model

vehicles. Traffic lights optimization means that different green and red light duration will be assigned to different streets in an intersection by intelligent traffic light instead of fixed and predefined duration. Optimal path is used for a path with low traffic and reasonable distance to destination in this paper. This model is illustrated in Figure 1 and is called Wings because of its similarity to real wings.

Referring to Figure 1, wireless devices are mounted on traffic light and vehicles. Consequently, vehicles inside the communication range of traffic lights, will send a message to their front traffic light. This message contains vehicles location, speed, direction and Received Signal Strength (RSS). Traffic lights will determine the farthest vehicle in each street based on their locations and RSS, and send request message to them. These farthest vehicles broadcast a request message in their communication rang in order to response to traffic lights request message. Vehicles located behind the farthest vehicle in this communication range will response to this message. Number of received messages will be sent to traffic light through the farthest vehicle. Thus, traffic light can calculate the number of near vehicles to intersection on ith street (N_{vi}) and total number of near vehicles to intersection (T_v) based on real-time data. A fixed and constant cycle length (C_l) is assigned to each traffic light by considering different parameters such as number of lanes, width of the streets, downtown vs. suburban streets and intersections and etc. Using this information, traffic light will assign different green light time durations (G_{ti}) to different streets in its communication range. These times will be calculated for each street by below formula:

$$G_{ti} = \frac{N_{vi}}{T_v} * C_l \tag{1}$$

Table 1 shows an example of green light time duration calculation for an intersection. Traffic light cycle length is assumed as 4 minutes.

Table 1. An example of green time duration calculation by traffic light

Road ID	Number of Vehicles	Assigned green time duration
A	14	63 sec
B	20	90 sec
C	2	9 sec
D	44	198 sec

3.2 Vehicular Routing with Optimal Path

As discussed in introduction section, through the past decade the number of vehicles grows sharply and cause many problems such as vehicles traffic and accident, long travel time, high CO2 and greenhouse gases emissions and fuel consumption. Building new high capacity streets and highways can alleviate some of aforementioned problems. However, this solution is very costly and most of the cases are not possible because of space limitations. Using Vehicle Route Guidance System (VRGS) is another way to utilize the roads capacity efficiently by proposing source-to-destination paths to drivers considering different objectives such as shortest or toll-free paths. But, most of the available navigators are using static routing algorithm such as Dijkstra or A* algorithms or in the best case are using dynamic traffic information (like TMC) but with rather high update intervals of several minutes. Nevertheless, these approaches need centralized process unit to compute the best or shortest path which limits the covered area and need high map update intervals in the system.

Thus, we propose a new decentralized routing algorithm based on real-time traffic information using vehicular networks and ant colony algorithm. Since, the vehicles traffic is the main source of problems in vehicle management systems based on RACQ Congested Roads report in [1]. Thus, our proposed algorithm is aiming to reduce the traffic in order to increase throughput while avoid to create another bottleneck at other street. Because congestion condition in vehicular networks is very dynamic and change as time goes by. Thus, we divide routing map into different cells and routing will be done based on current traffic condition on each cell. Moreover, layered model is used in order to reduce the computing overhead as well as increase the coverage area. Our layered and cellular model is illustrated in Figure 2 and is explained as follow:

Referring to Figure 2, our proposed model contains three different bottom-up layers:

1. Physical layer: This layer shows the real road map, nodes correspond to intersections, junctions, meanwhile, links correspond to streets and highways. This map can be exported from map databases like OpenStreetMap. This layer will be used for intra-cell (inside one cell) routing in our algorithm. This layers graph is given by $G_p = (N_p, L_p)$, where Np and L_p is the set of nodes and links, respectively. At each specific time (ti), a weight will be assigned to each link in the graph based on vehicles density ($NV_{ij}(_{ti})$). Vehicles density can be obtained from different tools such as Road Side Units

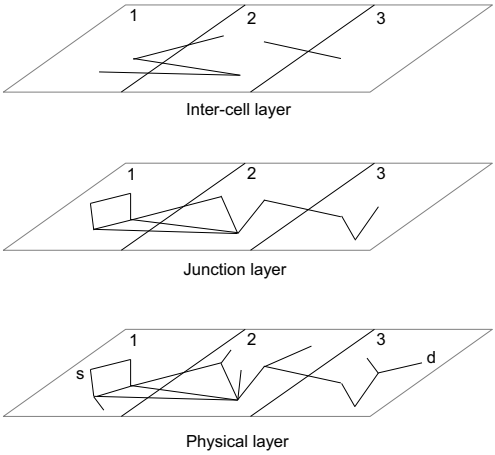

Fig. 2. Proposed layered and cellular model used in ant-based vehicle routing algorithm

(RSUs), Inductive Loop Detectors (ILD) [21] and Video Imaging Vehicle Detection System (VIVDS) [22]. In this paper, we assumed that each streets density is available through one of above mentioned ways. α_{ij} presents link weight between nodes i and j. In this paper, number of vehicles and links weight has inverse relationship, thus, α_{ij} can be calculated as follow:

$$\alpha_{ij} = \frac{1}{NV_{ij(ti)}} \tag{2}$$

2. Junction layer: In this layer, irrelevant nodes in physical layer which dont represent a junction are pruned.
3. Inter-cell layer: Junctions and their links which connect two different cells in junction layer will remain, otherwise, they are pruned. The remained nodes (junctions) are called border nodes. This layers information will be used whenever a vehicle travels over larger distances and thus traverses more than one cell to reach its destination. Inter-cell (between two different cells) routing table will be created based on this layers information. Table 2 is an example of inter-cell routing table for cell 1 of inter-cell layer illustrated in Figure 2.

Table 2. Inter-cell routing table for cell 1

Source	Destination	Vehicle Density
A1	A2	$NV_{A1,A2(ti)}$
A1	B2	$NV_{A1,B2(ti)}$
B1	B2	$NV_{B1,B2(ti)}$

In these tables, first and second columns show the existing path(s) between two different cells, and last column indicates vehicle density at particular time and thus it will be changes as time goes by based on number of vehicles on that path. Each cells inter-cell routing table will be disseminated among all junctions of same cell.

For example, Table 2 means that there are 3 outgoing links from 1st cell through two border nodes (A1, B1) to 2nd cell. Consequently, if a vehicle locates in cell 1 and want to travel to other cells (such as cell 2 or 3), first will be guided to one of these border junctions based on traffic condition using ACO algorithm, then based on traffic condition will be routed to one of border nodes in cell 2 using inter-cell routing table. If there are two or more path between two different cells (e.g. our example), path will lowest traffic will be selected.Therefore, vehicles will be routed through shortest low traffic paths, since researchers in [23] have been proven that ants find the shortest path.

Our last topic in this section is related to intra-cell routing process using ant-based agents. This process is based on ants behavior discussed in section 2. Our proposed algorithm contains three main steps:

1. Initialization: the pheromone values (weights), α_{ij}, on each link (path) are set based on vehicles density.
2. Pheromone Update: ants start to discover the rout between source and destination, and move to one of neighbor nodes based on pheromone values. This value will be decreased in two ways: first, over time by a factor ε using formula (3) and second, whenever an ant agent pass the link for finding a rout from source to destination by a factor β until a stop criterion (reach to destination or MAX-HOPS) is met by using formula (4).

$$\alpha_{ij}(t+1) = \alpha_{ij}(t) - \varepsilon \tag{3}$$

$$\alpha_{ij}(new) = (1-\beta) * \alpha_{ij}(current) \tag{4}$$

   This decreasing is done due to improve the exploration factor of the search. Because in this way, more new routes different from previous ones will be discovered and they can be used for traffic congestion mitigation purposes. MAX-HOPS is a constant value used for limiting the ants movements (e.g. Time To Live (TTL) value for routing packets).
3. Solution Construction: in this step the pheromone value will be increased only when an ant reaches the destination before it reaches MAX-HOPS. Ant backtracks to increase the pheromone levels on the links in found path by factor δ using formula (5).

$$\alpha_{ij}(new) = \alpha_{ij}(current) - \delta \tag{5}$$

In most other approaches, decreasing and increasing of pheromone values happen globally which requires synchronization and more communication. However, these updates are happened locally in our proposed method.

4 Simulation Results and Discussion

In order to evaluate our proposed approach for traffic lights optimization, DIVERT simulator [24] is used. A road topology with two intersections and bidirectional streets is used for evaluation. 100 vehicles with various speeds ranging from 40 km/h to 90 km/h, are distributed randomly in this topology. These vehicles moves toward an specific predefined points based on their directions Two scenarios, one with usual traffic lights and another with our proposed adaptive and dynamic traffic lights, were considered in this simulation. Vehicles average speed and waiting time as well as number of stopped vehicles are three evaluation metrics in our simulation. The number of stopped vehicles behind the traffic lights at intersections is demonstrated in Figure 3. Based in this figure during the simulation time, the number of stopped vehicles in adaptive and dynamic traffic lights (our proposed method) is less than the number of stopped vehicles

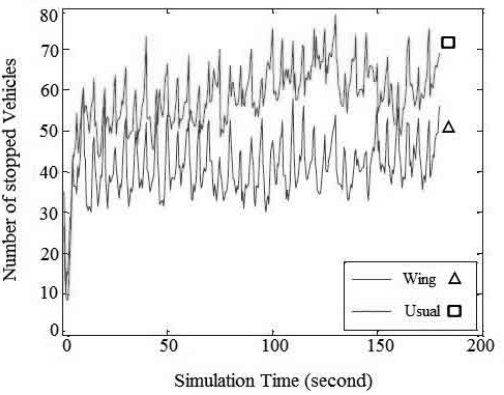

Fig. 3. Number of stopped vehicles for two scenarios

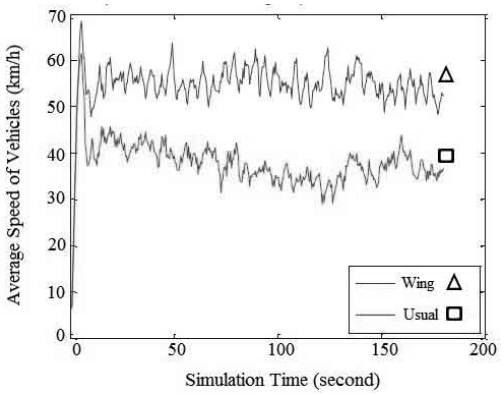

Fig. 4. Average speed of vehicles in two scenarios

in usual traffic light scenario. It means that traffic congestion at intersections are reduced using our method. Figure 4 compares the vehicles average speed in two aforementioned scenarios. Referring to Figure 4, vehicles average speed is higher in the case of using our proposed method because of lower number of stops and longer green times for high traffic streets.

5 Conclusion and Future Work

In this paper, we addressed one the most important problems in transportation system which is vehicle traffic congestion problem. Based on our literature review, traffic lights optimization and dynamic vehicle routing are two main approaches used for solving traffic congestion problem. Therefore, a cellular ant-based algorithm using optimized traffic lights which combines these two methods is proposed for traffic congestion problem in vehicular environments. Ant-based algorithm is used due to their superior ability in solving dynamic problems. Moreover, our traffic light optimization is done without using road side units which reduces the whole cost of the approach. This optimization examined through a simulation and results indicate that vehicles average speed and number of stopped vehicles at intersections are improved significantly as compared with usual traffic lights. As future work, we plan to implement the second part of our approach which is cellular ant-based algorithm for vehicle routing through shortest path with less traffic and compare it with static (e.g. Dijkstra) and dynamic (e.g. Dynamic System for the Avoidance of Traffic Jams (DSATJ)) approaches.

Acknowledgment. This research is supported by the Ministry of Higher Education (MOHE) and University of Malaya (project no. HIR-MOHEB00009).

References

1. Spalding, S.: Racq congested roads report (2008)
2. Kroon, R., Rothkrantz, L.: Dynamic vehicle routing using an abc-algorithm. Collected Papers on the PITA Project, 21 (2003)
3. Suson, A.C.: Dynamic routing using ant-based control
4. Tatomir, B., Rothkrantz, L.: Hierarchical routing in traffic using swarm-intelligence. In: Intelligent Transportation Systems Conference, ITSC 2006, pp. 230–235. IEEE (2006)
5. Kponyo, J.J., Kuang, Y., Li, Z.: Real time status collection and dynamic vehicular traffic control using ant colony optimization. In: 2012 International Conference on Computational Problem-Solving (ICCP), pp. 69–72. IEEE (2012)
6. Bonabeau, E., Dorigo, M., Theraulaz, G.: Swarm intelligence: from natural to artificial systems. Number 1. OUP USA (1999)
7. Dhillon, S., Van Mieghem, P.: Performance analysis of the antnet algorithm. Computer Networks 51(8), 2104–2125 (2007)
8. Di Caro, G., Dorigo, M.: Antnet: Distributed stigmergetic control for communications networks. arXiv preprint arXiv:1105.5449 (2011)

9. Stevanovic, A.: Adaptive traffic control systems: domestic and foreign state of practice. Number Project 20-5, Topic 40-03 (2010)
10. Akcelik, R., Besley, M., Chung, E.: An evaluation of scats master isolated control. In: Proceedings of the 19th ARRB Transport Research Conference (Transport 1998), pp. 1–24 (1998)
11. Robertson, D.I., Bretherton, R.D.: Optimizing networks of traffic signals in real time-the scoot method. IEEE Transactions on Vehicular Technology 40(1), 11–15 (1991)
12. Gradinescu, V., Gorgorin, C., Diaconescu, R., Cristea, V., Iftode, L.: Adaptive traffic lights using car-to-car communication. In: IEEE 65th Vehicular Technology Conference, VTC 2007-Spring, pp. 21–25. IEEE (2007)
13. Zhou, B., Cao, J., Zeng, X., Wu, H.: Adaptive traffic light control in wireless sensor network-based intelligent transportation system. In: 2010 IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall), pp. 1–5. IEEE (2010)
14. Faye, S., Chaudet, C., Demeure, I.: A distributed algorithm for multiple intersections adaptive traffic lights control using a wireless sensor networks. In: Proceedings of the First Workshop on Urban Networking, pp. 13–18. ACM (2012)
15. Ferreira, M., Fernandes, R., Conceição, H., Viriyasitavat, W., Tonguz, O.K.: Self-organized traffic control. In: Proceedings of the seventh ACM International Workshop on VehiculAr InterNETworking, pp. 85–90. ACM (2010)
16. Kassabalidis, I., El-Sharkawi, M., Marks, R., Arabshahi, P., Gray, A., et al.: Adaptive-sdr: Adaptive swarm-based distributed routing. In: Proceedings of the 2002 International Joint Conference on Neural Networks, IJCNN 2002, vol. 1, pp. 351–354. IEEE (2002)
17. Soltani, A., Akbarzadeh-T, M.R., Naghibzadeh, M.: Helping ants for adaptive network routing. Journal of the Franklin Institute 343(4), 389–403 (2006)
18. Tekiner, F., Ghassemlooy, Z., Al-khayatt, S.: Antnet routing algorithm-improved version. In: CSNDSP 2004, Newcastle, UK, pp. 22–22 (2004)
19. Tatomir, B., Rothkrantz, L.: Dynamic traffic routing using ant based control. In: 2004 IEEE International Conference on Systems, Man and Cybernetics, vol. 4, pp. 3970–3975. IEEE (2004)
20. Claes, R., Holvoet, T.: Ant colony optimization applied to route planning using link travel time predictions. In: 2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), pp. 358–365. IEEE (2011)
21. Roland, N.: Inductive loop detector. US Patent 20,050,035,880 (February 17, 2005)
22. Michalopoulos, P.G.: Vehicle detection video through image processing: the autoscope system. IEEE Transactions on Vehicular Technology 40(1), 21–29 (1991)
23. Jayadeva, Shah, S., Bhaya, A., Kothari, R., Chandra, S.: Ants find the shortest path: a mathematical proof. Swarm Intelligence 7(1), 43–62 (2013), doi: 10.1007/s11721-013-0076-9
24. LIACC: Divert: Development of inter-vehicular reliable telematics (2013)

Robust Inventory Management under Uncertain Demand and Unreliable Delivery Channels

Przemysław Ignaciuk

Institute of Information Technology, Lodz University of Technology,
215 Wólczańska St., 90-924 Łódź
przemyslaw.ignaciuk@p.lodz.pl

Abstract. The paper considers the problem of establishing an efficient supply policy for production-inventory systems in which stock replenishment process is unreliable. In the analyzed setting, the stock at a goods distribution center, used to satisfy uncertain, variable demand is refilled using multiple delivery channels. Due to information distortion, product defects, or improper transportation, the shipments may arrive damaged, or incomplete. The setting is modeled as a time-varying discrete-time system with multiple input-output delays. A new delay compensation mechanism, which provides smooth ordering pattern and ensures closed-loop stability for arbitrary delay, is proposed. Conditions for achieving full satisfaction of the *a priori* unknown demand are specified and formally proved.

Keywords: inventory control, time-delay systems, discrete-time systems.

1 Introduction

It has been argued in a number of recent works [1, 2] that in the currently observed increased competition and demand diversity one might seek performance improvements in production-inventory systems through the application of systematic design techniques. Therefore, as opposed to the classical stochastic, or heuristic solutions to inventory control problem, in this work formal approach is adopted.

In the considered class of systems the stock accumulated at a goods distribution center is used to satisfy uncertain market demand. Neither the value, nor statistics of demand are known *a priori*, and thus it is treated as a disturbance. The stock is replenished with delay using (possibly) multiple supply sources [3], or delivery channels [4]. Unlike the previous studies in a similar setting with nonnegligible delay and uncertain demand [5]–[11], in this paper, the situation when the stock replenishment process itself is subject to perturbation is analyzed. The investigated extra source of uncertainty is related to information distortion (e.g. erroneous order handling) and faults in goods production and delivery. In order to perform sound, formal analysis, a new model of the considered class of production-inventory systems is constructed. In the proposed framework, the uncertainty related to unreliable delivery channels is modeled as an external multiplicative disturbance. A new control strategy (ordering

J. Świątek et al. (eds.), *Advances in Systems Science*,
Advances in Intelligent Systems and Computing 240,
DOI: 10.1007/978-3-319-01857-7_7,

policy), explicitly taking into account the problems related to unreliable replenishment process, is proposed. A pivotal role in establishing robust yet efficient ordering pattern plays a new delay compensation mechanism incorporated in the proposed control scheme. The designed compensation mechanism allows one to maintain closed-loop stability without compromising response speed, which is difficult to achieve in time-delay systems subject to perturbations [12, 13].

It is shown that the order quantities generated by the proposed strategy are always nonnegative and bounded, which is required for the practical implementation of any efficient ordering policy. It is also demonstrated that in the analyzed control system the available stock is never entirely depleted despite unpredictable demand variations and delivery channel uncertainty. As a result, all of the imposed demand can be satisfied from the readily available resources and maximum service level is obtained. Moreover, the storage space to accommodate all the incoming shipments is indicated, which helps in establishing suitable warehousing solutions. The crucial system properties are strictly proved and illustrated with numerical data.

2 System Model

The model of the analyzed production-inventory system is illustrated in Fig. 1. The system variables are inspected at regular, discrete time instants kT, where T is the review period and $k = 0, 1, 2, \ldots$ In order to save on notation in the remainder of the paper k will be used as the independent variable in place of kT.

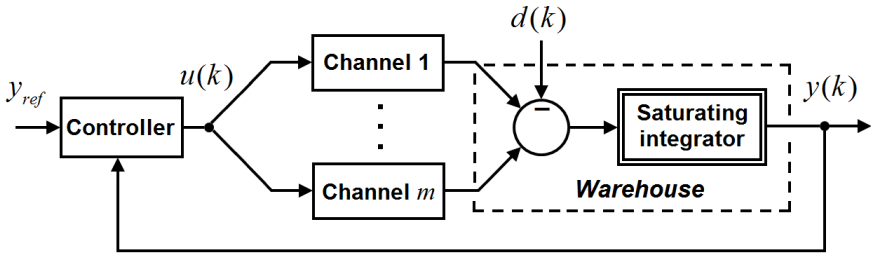

Fig. 1. Model of production-inventory system with unreliable delivery channels

The imposed demand (the goods quantity requested from inventory in period k) is modeled as an *a priori* unknown, bounded function of time $d(k)$, $0 \le d(k) \le d_{\max}$, where $d_{\max}$ is a positive constant denoting the estimate of maximum demand. From the control system perspective, the demand, being an exogenous, uncertain signal, is treated as a disturbance. If there is sufficient amount of goods in the warehouse to satisfy the imposed demand $d(k)$, then the actually met demand $h(k)$ (the goods sold to the customers or sent to the retailers in a distribution network) will be equal to the requested one. Otherwise, the imposed demand is satisfied only from the arriving shipments and additional demand is lost (it is assumed that the sales are not backordered and the excessive demand is equivalent to a missed business opportunity). Thus, one may write