Revista Brasileira de Ciências Mecânicas

Journal of the Brazilian Society of Mechanical Sciences

2

PUBLICAÇÃO DA ABCM - ASSOCIAÇÃO BRASILEIRA DE CIÊNCIAS MECÂNICAS

VOL. XIX - No. 2 - JUNE 1997 ISSN 0100-7386

JOURNAL OF THE BRAZILIAN SOCIETY OF MECHANICAL SCIENCES
REVISTA BRASILEIRA DE CIÊNCIAS MECÂNICAS

REVISTA BRASILEIRA DE CIÊNCIAS MECÂNICAS
JOURNAL OF THE BRAZILIAN SOCIETY OF MECHANICAL SCIENCES
Vol. 1, Nº 1 (1979)-
Rio de Janeiro: Associação Brasileira de Ciências Mecânicas
Trimestral
Inclui referências bibliográficas.
1. Mecânica
ISSN-0100-7386

A REVISTA BRASILEIRA DE CIÊNCIAS MECÂNICA publica trabalhos que cobrem os vários aspectos da ciência e da tecnologia em Engenharia Mecânica, incluindo interfaces com as Engenharias Civil, Elétrica, Química, Naval, Nuclear, Aeroespacial Alimentos, Agrícola, Petróleo, Materiais, etc., bem como aplicações da Física e da Matemática à Mecânica.

INDEXED by Applied Mechanics Reviews and Engineering Information, Inc.

Publicação da / Published by
ASSOCIAÇÃO BRASILEIRA DE CIÊNCIAS MECÂNICAS
THE BRAZILIAN SOCIETY OF MECHANICAL SCIENCES

Secretária da ABCM : **Ana LÚcia Fróes de Souza**
Avenida Rio Branco, 124 18º Andar
20040-001 Rio de Janeiro RJ
Tel.: (021) 221-0438/Fax: (021) 222-7128

Presidente: **Carlos Alberto de Almeida**
Vice-Presidente: **Arthur Palmeira Ripper Neto**
Secretário Geral: **Hans Ingo Weber**
Diretor de Patrimônio: **Felipe Bastos de F. Rachid**
Secretário: **Paulo Batista Gonçalves**

Secretária da RBCM: **Maria de Fátima Alonso de Souza**
UNICAMP - FEM - C.P. 6122
13083-970 Campinas SP
Tel.: (019) 239-8353/Fax: (019) 239-3722

REVISTA FINANCIADA COM RECURSOS DO
Programa de Apoio a Publicações Científicas
MCT CNPq FINEP

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 108-120

ISSN 0100-7386
Printed in Brazil

Cascade Control of Hydraulic Actuators

Raul Guenther
UFSC - Universidade Federal de Santa Catarina
Departamento de Engenharia Mecânica - C.P. 476

Edson Roberto De Pieri
Departamento de Engenharia Elétrica - C.P. 476
88049-900 Florianópolis, SC Brasil

Abstract

In this paper the problem of hydraulic positioning system is taken into account. It is shown that by modeling the actuator in a convenient form the system can be interpreted as a mechanical subsystem driven by a hydraulic one. This structure is called a cascade control strategy. Cascade controllers based on the linear an nonlinear models are synthesized using Lyapunov direct method. Simulation results illustrated the main characteristics of the cascade control strategy.
Keywords: *Hydraulic Control Systems, Hydraulic Actuators, Positioning Systems.*

Introduction

In many robot applications where big loads have to be handled hydraulic actuators are very attractive due to their ability to provide very high forces (and torques) at very high power levels. This results in higher torque/mass ratios than those available from an equivalent electric actuator. Another advantage of a hydraulic actuator is that it is quite stiff when viewed from the side of the load. This is because a hydraulic medium is mechanically stiffer than an electromagnetic one.

However, the compressibility of the hydraulic fluid can cause detrimental effects on the dynamic response of the hydraulic actuators, including instability and limited bandwidth. These effects limits the use of hydraulic actuators, specially in robot applications where high performance and reduced position errors are required.

In this paper we propose a control algorithm to a hydraulic positioning system, in order to overcome the limitation in the use of hydraulic actuators. Based on the methodology of order reduction presented in Utkin (1987), we propose a cascade control strategy.

The cascade control is based on a standard linear third order model, obtained from the linearization of the nonlinear differential equations. This mathematical model represents the dynamical behavior of hydraulic actuators.

By combining the cascade control scheme and the Lyapunov stability method we demonstrate that the closed loop error is globally exponential stable. This allows the improvement of the dynamic performance of the closed loop system. This result is not verified when classical controllers such as the PID (Proportional-Integral-Derivative) are used.

The dynamic performance of classical controllers and the proposed cascade control scheme are compared on a benchmark example. Simulation results based on the linearized model confirm the accuracy of the cascade control strategy. These results show that the cascade control is a powerful method to obtain high performance with hydraulic actuators.

The paper is organized as follows. First we present the nonlinear and the standard linear third order model. Then we discuss the hydraulic actuator control problem. The cascade control strategy applied to the hydraulic actuator is then presented. Next we demonstrate the cascade control stability properties. Finally, some simulations results for the benchmark hydraulic actuator (Davidson, 1990) are presented to illustrated the performance of the proposed controller.

The Dynamic Model

We consider the hydraulic actuator shown in Fig. 1. This actuator consists of a cylinder controlled by a servovalve constituted of a critical center four-way spool (Watton, 1989). The supply fluid pressure p_s is used for the energy transmission.

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

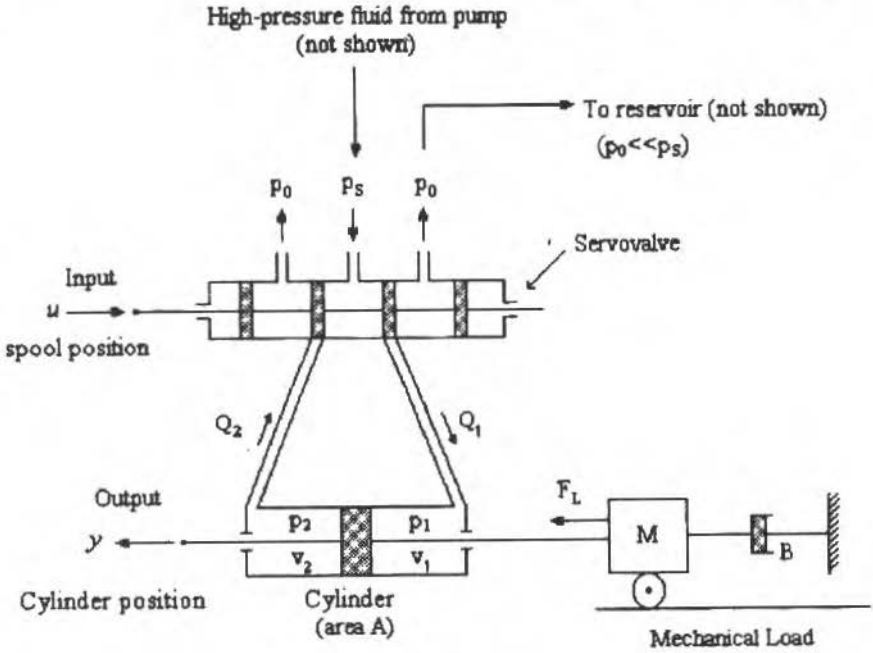


Fig. 1 Hydraulic Actuator

In Figure 1, p_0 is the reservoir pressure, p_1 is the pressure in the 1st cylinder chamber, p_2 is the pressure in the 2nd cylinder chamber, v_1 is the volume of chamber 1, v_2 is the volume of chamber 2, Q_1 is the flow through the servovalve from the pump to the chamber 1, Q_2 is the output flow from the chamber 2 to the reservoir, M is the mass attached to the actuator, B is the viscous friction coefficient, "u" is the servovalve spool position, and "y" is the actuator piston position (and consequently the position of the mass M).

The dynamic model of this system is obtained using the well-known flow equation. This modeling is developed in terms of energy conservation and applying the continuity equation to each cylinder chamber. The final model is obtained applying he Newton second law to the mechanical mass damper system relating the input "u" (spool position) to the output "y" (cylinder position). This modeling process is developed by many authors (Stringer 1976, Watton 1989, Burton 1994), and detailed information can be found in Paim (1996), where the following description is obtained:

$$M\ddot{y} + B\dot{y} = Ap_{\Delta} + F_L \tag{1}$$

$$\dot{p}_{\Delta} + \frac{\beta v}{(v/2)^2 - (Ay)} A\dot{y} = \frac{\beta v}{(v/2)^2 - (Ay)} Ku\sqrt{p_s - sign(u)p_{\Delta}} \tag{2}$$

where A is the cross-sectional area of the cylinder piston, $p_{\Delta} = p_1 - p_2$ is the pressure difference in the cylinder, β is the volumetric elasticity modulus, $v = v_1 + v_2$ is the total volume of the cylinder, and F_L is the load applied to the mass M. The other variables have been defined above.

By considering the case $F_L = 0$ (the load will be considered latter as an external disturbance), the linearization of the model defined by Eqs. (1) and (2) around the central position of the actuator (y=0) allows to write (Watton 1989):

$$M\ddot{y} + B\dot{y} = Ap_{\Delta} \tag{3}$$

$$\frac{v}{4\beta K_Q}\dot{p}_{\Delta} + \frac{K_C}{K_Q}p_{\Delta} + \frac{A}{K_Q}\dot{y} = x \tag{4}$$

where K_Q is the valve flow coefficient, K_C is the valve pressure coefficient.

According to Burton (1994), we can define a state vector as $x = \begin{bmatrix} y & \dot{y} & p_{\Delta} \end{bmatrix}^T$, then Eqs. (3) and (4) can be written in the state space representation as:

$$\dot{x} = Fx + Gu$$

$$y = Hx \tag{5}$$

where $F = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -\frac{B}{M} & \frac{A}{M} \\ 0 & -\frac{4\beta A}{v} & -\frac{4\beta K_C}{v} \end{bmatrix}$ $G = \begin{bmatrix} 0 \\ 0 \\ \frac{4\beta K_Q}{v} \end{bmatrix}$ $H = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$

By combining Eqs. (3) and (4) we obtain a third order transfer function relating the input "u" to the output "y" of the dynamic system:

$$G(s) = \frac{4\beta}{v}\frac{A}{M}K_Q \frac{1}{s\left(s^2 + 2\zeta\omega_n + \omega_n^2\right)} \tag{6}$$

where the natural frequency ω_n is given by $\omega_n = 2\sqrt{\frac{\beta}{Mv}(A^2 + BK_C)}$.

It should be remarked that:

i - The linear model of the hydraulic actuator is represented by a transfer function with a single pole in the origin and two complex conjugated poles poorly damped, leading to a lightly damped system response;

ii - The natural frequency of these complex pole ω_n depends on the volumetric elasticity modulus of the fluid and the actuator position. Particularly, the central position leads to the minimal natural frequency in relation to the other positions (Watton 1989);

iii - In this model the servovalve is considered as a static component. In fact, the transfer function associated to the valve has real and complex poles superimposed to that of the transfer function $G(s)_{u \to y}$ described in (6) (Schothorst et al., 1994). In Heintze and Weiden (1995) servovalve dynamics is represented by a third order linear system, and

iv - When the pipeline between the servovalve and the hydraulic cylinder is long it can cause a dynamic behavior in this pipeline. This phenomenon is called pipeline dynamics. These dynamics (see Schothorst et al., 1994) can be represented as second order systems with lightly damped poles in series with the servovalve dynamics, and with the subsystem represented by the transfer function (6).

The Hydraulic Actuator Control

Due the particular structure of the system represented by Eqs. (3) and (4) we can easily evaluate the limitations of the closed loop system performance. The presence of lightly damped complex conjugated poles implies that the proportional feedback gain may be restricted to small values. In fact, great values of the gain lead to locate the closed loop poles in the right side of the complex plane. This restriction in the gain margin is a hard limitation to the system performance. Therefore, the position errors become difficult to be minimized.

The presence of lightly damped poles lead the system to be poorly robust with respect to external disturbances like the load F_L, and the non modeled dynamics, e.g. the servovalve and the pipeline dynamics.

The PD (Proportional-Derivative), PI (Proportional-Integral), or PID (Proportional-Integral-Derivative) controllers introduce poles and zeros in the closed loop system that, in general, do not change the locations of the complex conjugated poles. This feature can be verified if we plot the corresponding root locus for each controller. Consequently, with the output feedback using these controllers is quite difficult to change the bandwidth, or to improve the system performance, or yet to reduce the position errors.

The damping coefficient of the complex conjugated poles associated with the hydraulic components has been recognized as an essential topic of investigations by many authors. Watton (1989) presents four possibilities to improve the system damping: (i) Introducing a leakage way between the lines with pressure p_1 and p_2, (ii) Using a hydromechanical filter between the lines, (iii) Employing an electronic filter, (iv) Using a state feedback strategy. These possibilities are discussed in the sequence.

The introduction of a leakage way between the lines can improve the damping as demonstrated by simulations presented in Watton (1989). However, to an effective damping it is necessary to limit the closed loop proportional gain. This limitation can be expected to play an important role in the possibility of reducing the position errors and the bandwidth.

The main feature of a hydromechanical filter is to lead to a leakage between the lines only in the high frequencies corresponding to the lightly damped poles. Consequently, it is possible to reduce the position error without substantial changing in the bandwidth as in the case of the direct leakage between the lines.

Employing an electronic filter with restricted band is a way to avoid the displacement of the closed loop poles to the right side of the complex plane. The basic idea is to introduce two zeros in a way to cancel the complex conjugate poles. There are two difficulties to the practical implementation of this technique. Firstly, it is imperative that the zeros velocities be smaller than the natural frequencies of the poles. If this is not the case, then the introduced zeros are not able to maintain the poles in the left half plane (see Franklin et al., 1994). It is important to note that the natural frequency associated with the poles depends on the volumetric elasticity modulus β and the cylinder position. As the volumetric elasticity modulus β and the cylinder position vary, we conclude that, in the project, a small error in estimating the natural frequency of the poles could lead to closed loop instabilities.

The second difficulty is related to the filter. In case of practical implementation, the knowledge about the actuator acceleration $\ddot{y}(t)$ and its time derivative are required.

Furthermore, as the approximated cancellation of the poles does not eliminate the lightly damped dynamics, then the gains, bandwidth and performance remain limited.

Another way to increase the damping of the complex conjugated poles in the closed loop, is to synthesize a state feedback based controller as presented in Virtanen (1993). In theory, an arbitrary pole placement is always possible due to the fact that the system (3) (4) is completely controllable. However, a suitable damping of the complex poles can lead to high control gains. Due to the physical limitations, the servovalve may be unable to supply the needs of the pressure transfer, and a saturation can occurs limiting the system performance.

In the next section we will present a way to keep into account these limitations imposed by the problem structure given in Eqs. (3) and (4). The main objective is to obtain high performance

requirements where an accurately tracking is necessary, such as in case of hydraulic robot manipulators.

The Cascade Control Strategy

Based on the methodology of order reduction presented in Utkin (1987), we propose a cascade control strategy described as follows. Consider a system of the form

$$\dot{x}_1 = h_1(x_1, x_2, t) \tag{7}$$

$$\dot{x}_2 = h_2(x_1, x_2, t) + B_2(x_1, x_2, t)u \tag{8}$$

where $x_1 \in \Re^{n-m}$, $x_2 \in \Re^m$ and $u \in \Re^m$, "n" is the order of the subsystem modeled by (7) and (8) and "m" is the order of the subsystem (8). Consider that in system (7) (8) the vector $x_1(t)$ is required to track some desired trajectory $x_{1d}(t)$.

The cascade control design consists in regarding the vector x_2 as control variable for subsystem (7) and, as such, a desired function x_{2d}, to be designed, is assumed to exist so that the tracking goal may be achieved. Then $u \in \Re^m$ is designed so that x_2 tracks x_{2d}. In turn, this allows x_1 to track the desired trajectory $x_{1d}(t)$.

An application of this cascade strategy has been used by Guenther and Hsu (1993) in the control of robot manipulators with electric actuators, and by Hsu and Guenther (1993) to control flexible joint manipulators. Recently a kind of cascade control has been used by Heintze and Weiden (1995) to control hydraulic manipulators.

Inspecting Eqs. (3) and (4) we can rewrite them in the form (7) and (8). In fact, the system (3) (4) can be interpreted as a mechanical subsystem (3) driven by an hydraulic force $g = Ap_\Delta$, on which a hydraulic subsystem is superimpose to provide a pressure drop p_Δ, when driven by the spool displacement "u". This interpretation enforce the cascade model (3) (4) description.

To describe (3) (4) as a cascade system, we will define $g_d = p_{\Delta d}A$ as a desired force so that the mass M achieve a trajectory $y_d(t)$. Let

$$\tilde{p}_\Delta = p_\Delta - p_{\Delta d}$$

be the pressure difference tracking error. Using (9) we can rewrite Eqs. (3) and (4) as:

$$M\ddot{y} + B\dot{y} = Ap_{\Delta d} + A\tilde{p}_\Delta \tag{10}$$

$$\frac{v}{4\beta K_Q}\dot{p}_\Delta + \frac{K_C}{K_Q}p_\Delta + \frac{A}{K_Q}\dot{y} = u \tag{11}$$

Clearly, this system is in the cascade form as stated in (8) and (9). Equation (10) can be interpreted as a mechanical second order subsystem actuated with a desired force g_d subjected to an input perturbation $d = A\tilde{p}_\Delta$. Equation (11) represents the hydraulic subsystem.

The design of the cascade controller can be summarized as follows:

(i) Compute a control law g_d to he mechanical subsystem (10) such that the cylinder displacement achieve a desired trajectory $y_d(t)$ taking into account the presence of the perturbation $d = A\tilde{p}_\Delta$. We can quantify the desired pressure difference using the relation:

$$p_{\Delta d} = \frac{g_d}{A} \tag{12}$$

(ii) Compute a control law "u" such that p_Δ tracks $p_{\Delta d}(t)$ defined in (12).

In this paper the design of a control law to the mechanical subsystem g_d is based on the controller proposed by Slotine and Li (1988). The control law "u" is synthesized to achieve good performance characteristics to the tracking related to the hydraulic subsystem.

According to Slotine and Li (1988) the control law to obtain tracking in the mechanical subsystem is given by

$$g_d = M\ddot{y}_r + B\dot{y}_r - K_D s \tag{13}$$

where K_D is a positive constant, $\dot{y}_r$ is the reference velocity and s is a measure of the velocity tracking error. In fact, $\dot{y}_r$ can be obtained in modifying the desired velocity $\dot{y}_d$ as follows

$$\dot{y}_r = \dot{y}_d - \lambda\tilde{y} \; ; \qquad \tilde{y} = y - y_d \; ; \qquad s = \dot{y} - \dot{y}_r = \dot{\tilde{y}} + \lambda\tilde{y} \tag{14}$$

where λ is a positive constant.

Substituting (13) in (10) the error equation related to the mechanical subsystem becomes

$$M\dot{s} + (B + K_D)s = A\tilde{p}_\Delta \tag{15}$$

Consider the nonnegative function:

$$2V_1 = Ms^2 + P\tilde{y}^2 \tag{16}$$

where P is a positive constant defined in the sequel. Using (15), the time derivative of (16) is given by

$$\dot{V}_1 = -(B + K_D)s^2 + P\tilde{y}\dot{\tilde{y}} + As\tilde{p}_\Delta \tag{17}$$

Expression (17) will be on the stability analysis.

To achieve the trajectory tracking in the hydraulic subsystem (11) we propose the control law

$$u = \hat{u} - K_p\tilde{p}_\Delta \tag{18}$$

where K_p is a positive constant and $\hat{u}$ is a nominal control law. This nominal control law is designed to the system represented by its nominal parameter values.

The design of $\hat{u}$ and K_p is based on a nonnegative scalar function V_2:

$$2V_2 = L\tilde{p}_\Delta^2 \tag{19}$$

where $L = \dfrac{v}{4\beta K_Q}$ is a positive constant.

The time derivative of (19) using Eq. (11), and taking into account the pressure difference error $\tilde{p}_\Delta(t)$ and the control law defined in (18) results

$$\dot{V}_2 = -\tilde{p}_\Delta\left[(u^* - \hat{u}) + K_p\tilde{p}_\Delta\right] \tag{20}$$

where

$$u^* = L\dot{p}_{\Delta d} + \frac{K_C}{K_Q}p_\Delta + \frac{A}{K_Q}\dot{y} \tag{21}$$

Expression (20) will be used on the stability analysis.

Considering the cascade control scheme proposed, the spool displacement "u" can be computed using the relationship

$$u = \hat{u} - K_p(p_\Delta - p_{\Delta d}) \tag{22}$$

with the nominal control û given by

$$\hat{u} = \hat{L}\dot{p}_{\Delta d} + \frac{\hat{K}_C}{\hat{K}_Q}p_\Delta + \frac{\hat{A}}{\hat{K}_Q}\dot{y} \tag{23}$$

where $\hat{L} = \dfrac{\hat{v}}{4\hat{\beta}\hat{K}_Q}$ is a constant determined using the nominal values related to the volume $\hat{v}$, the volumetric elasticity modulus $\hat{\beta}$, the value flow coefficient $\hat{K}_Q$, the valve pressure coefficient $\hat{K}_C$ and the cross-sectional area of the actuator piston Â.

The desired pressure difference $p_{\Delta d}$ in Eq. (22) can be computed using Eqs. (12) and (13):

$$p_{\Delta d} = \frac{g_d}{A} = \frac{1}{A}(M\ddot{y}_r + B\dot{y}_r - K_D s) \tag{24}$$

The desired pressure difference time derivative $\dot{p}_{\Delta d}$ in Eq. (23) is the time derivative of (24) whose computation involves the relationships $\dfrac{d^3}{dt^3}(y_r) = y_r^{(3)}$ and $\dfrac{d}{dt}(s) = \dot{s}$. From equation (14)

we can conclude the necessity of the knowledge of the cylinder acceleration $\ddot{y}$. In the case that all parameters related to the mechanical subsystem are known, $\ddot{y}$ can be computed using Eq. (3). Using this approach and expression (14) it is possible to write:

$$\dot{p}_{\Delta d} = \frac{1}{A}\left[\left(\frac{K_D B}{M} - \lambda K_D\right)\dot{y} - A\left(\frac{K_D}{M} + \lambda\right)p_\Delta + My_d^{(3)} + (B + K_D + \lambda M)\ddot{y}_d + \lambda(B + K_D)\dot{y}_d\right] \tag{25}$$

We can note from (21) and (25) that the signal u* is obtained from measuring the variables y, $\dot{y}$ and p_Δ, i.e., the state vector "x" associated with the system (3) (4).

Stability Analysis

Combining the trajectory tracking, algorithms presented before a cascade controller can be obtained. Now consider the linear mathematical model of the hydraulic actuator and the cascade controller. In this case, the closed loop system is given by Ω = {(10),(11),(13),(18)}. Let $\rho = \begin{bmatrix} \tilde{y} & \dot{\tilde{y}} & \tilde{p}_\Delta \end{bmatrix}^T$ be the tracking error vector of Ω.

To demonstrate the exponential stability of the closed loop system Ω we will take into account the convergence lemma presented in the sequel.

Lemma - If a real valued function V(t) verify the inequality $\dot{V}(t)+\alpha V(t)\leq 0$ where α is a real number, then $V(t) \leq V(0)\, e^{-\alpha t}$ (see Slotine and Li, 1991, pp. 91).

Theorem - The Ω system above defined is exponentially stable with respect to the origin of the tracking error vector ρ.

Proof: Consider a Lyapunov function

$$V = V_1 + V_2 = \frac{1}{2}\rho^T N_1 \rho \tag{26}$$

where V_1 and V_2 have been defined in (16) and (19), respectively. The matrix N_1 is given by

$$N_1 = \begin{bmatrix} \lambda^2 M + P & \lambda M & 0 \\ \lambda M & M & 0 \\ 0 & 0 & L \end{bmatrix} \tag{27}$$

In the case of the parameters referring to the hydraulic actuator are known $\hat{u} = u^*$. Then, according to (17) and (20), the time derivative of (26) is

$$\dot{V} = -(B + K_D)s^2 + P\tilde{y}\dot{\tilde{y}} + As\tilde{p}_\Delta - K_p\tilde{p}_\Delta^2 \tag{28}$$

Using Eq. (14)

$$\dot{V} = -(B + K_D)\dot{\tilde{y}}^2 - \left[2\lambda(B + K_D) - P\right]\tilde{y}\dot{\tilde{y}} - \lambda^2(B + K_D)\tilde{y}^2 + \lambda A\tilde{y}\tilde{p}_\Delta + A\dot{\tilde{y}}\tilde{p}_\Delta - K_p\tilde{p}_\Delta^2 \tag{29}$$

Define $P = 2\lambda\left(B + K_D\right)$, then

$$\dot{V} = -\rho^T N_2 \rho \tag{30}$$

where

$$N_2 = \begin{bmatrix} \lambda^2(B + K_D) & 0 & -\frac{1}{2}\lambda A \\ 0 & B + K_D & -\frac{1}{2}A \\ -\frac{1}{2}\lambda A & -\frac{1}{2}A & K_p \end{bmatrix} \tag{31}$$

In terms of the controller parameters, if

$$K_D K_P > \frac{1}{2} A^2 \tag{32}$$

then the matrix N_2 is positive definite, and

$$\dot{V}(t) = -\rho^T N_2 \rho \leq 0 \tag{33}$$

Let $\lambda_{max}(N_1)$ be the maximum eigenvalue of the matrix N_1, and $\lambda_{min}(N_2)$ be the minimum eigenvalue of N_2, and let $\gamma = \lambda_{min}(N_2)/\lambda_{max}(N_1)$. As N_1 and N_2 are definite positive matrices, consequently all these scalars are positives. From matrix theory $N_1 \leq \lambda_{max}(N_1) I$ and $\lambda_{min}(N_2) I \leq (N_2)$, where I is the identity matrix (see Noble and Daniels 1986). Thus,

$$\rho^T N_2 \rho \geq \frac{\lambda_{min}(N_2)}{\lambda_{max}(N_1)} \rho^T \left[\lambda_{max}(N_1) I\right] \rho \geq \gamma V(t) \tag{34}$$

The use of this result combined with (33) enable us to write $\dot{V}(t) \leq -\gamma V(t)$. According to the convergence lemma, this result leads to

$$V(t) = \frac{1}{2} \rho^T N_1 \rho \leq V(0) e^{-\gamma t} \tag{35}$$

From (35) and using the relationship $\rho^T N_1 \rho \geq \lambda_{min}(N_1) \|\rho(t)\|^2$ we can state that the error vector associated with system ρ converges exponentially to the origin with a convergence rate greater than $\gamma/2$.

As stated by Vidyasagar and Vannelli (1982), the exponential convergence of the closed loop Ω implies that Ω is robust with respect to limited disturbances like the load F_L. Furthermore it has been demonstrated by Bodson and Sastry (1989) that exponential ensures robustness against nonmodeled dynamics like the servovalve and the pipeline dynamics in the hydraulic actuator.

It is clear that if expression (32) is verified then high gains can be used so that performance requirements can be achieved, i.e. we have no gain limitations. Consequently, the cascade controller overcomes limitations which inherently appear in the classical controllers. In fact, for practical purposes, gains are limited by non modeled dynamics concerning the servovalve and the pipelines.

Simulations Results

This example illustrates how to implement the cascade control strategy presented in section 5 for hydraulic actuator proposed as a benchmark in Davidson (1990).

The hydraulic actuator linear model is described by (3) (4). The benchmark data are (Davidson, 1990): M = 128.7 Kg; B = 162.73 Ns/m; A = 1.075 x 10^{-3} m^2; v = 0.874 x 10^{-3} m^3; β = 14 x 10^8 N/m^2. The servovalve parameters are (Paim, 1996): K_C = 2.5 x 10^{-12} m^5/Ns; K_Q = 0.76 m^2/s. The hydraulic cylinder corresponding to this benchmark data is 800 mm long.

In the cascade strategy, the control of the mechanical subsystem is designed according to (13) where K_D = 5000, and λ = 200.

The control law for the hydraulic subsystem is given by (18) where K_P = 1.0 x 10^{-9}, the nominal control $\hat{u}$ is given by (23) with $\hat{L} = L, \hat{K}_C = K_C, \hat{K}_Q = K_Q, \hat{A} = A$, and the desired pressure different derivative given by (25).

The hydraulic actuator closed loop performance with the cascade controller (CC) is compared with a proportional controller and a proportional derivative controller.

The proportional controller (P) is designed as $u_P = P(y - y_d)$, and the proportional derivative controller (PD) is $u_{PD} = P(y - y_d) + D(\dot{y} - \dot{y}_d)$.

We have employed two desired trajectories. A first one to illustrate our results in a regulation situation. In this case the desired trajectory starts in $y_d(0) = 0$ and achieve $y_d(1) = 100$ mm in 1 sec, according to a function described by the 7th order polynomial $y_d = -2t^7 + 7t^6 - 8{,}4t^5 + 3{,}5t^4$. A seconde one is used to verify the closed loop performance in a trajectory tracking situation. In this case the desired trajectory is described by $y_d(t) = 100 \sin(0.5\,\pi t)$.

Figure 2 shows the hydraulic actuator responses with the cascade controller (CC) and the proportional controller (P). The desired trajectory is described by the 7th order polynomial and the proportional gain is $P=4\times10^{-3}$, which corresponds to the limit of stability.

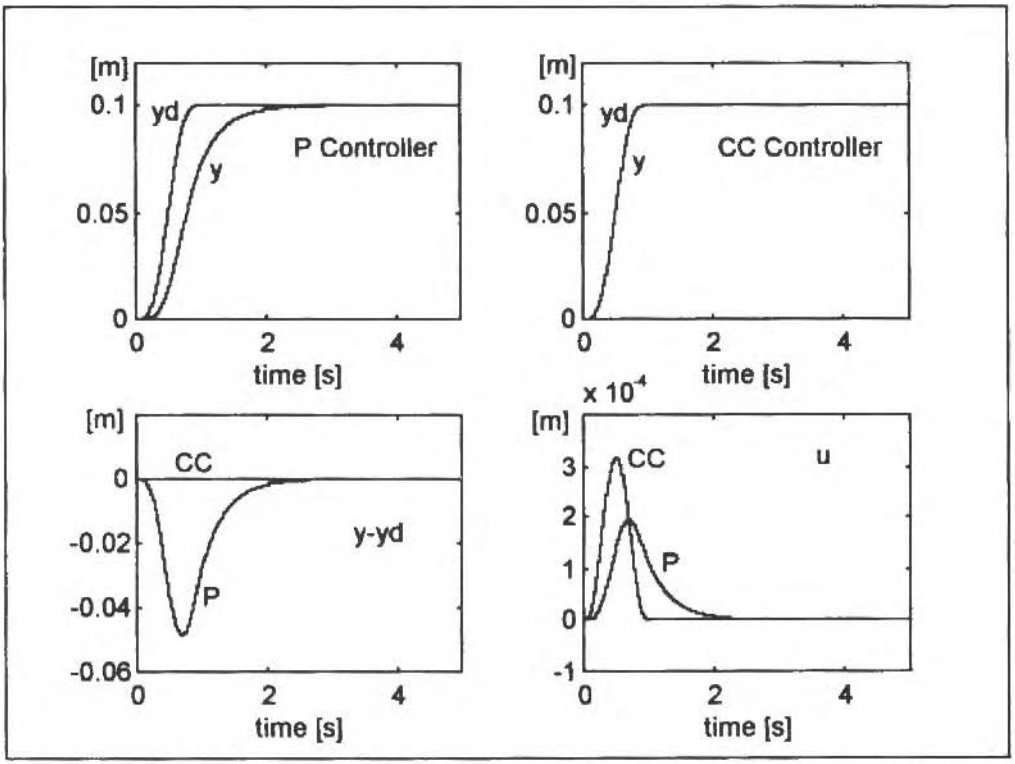


Fig. 2 Linear Model Responses to the P and CC Controller With a Desired Trajectory Described by the 7th Order Polynomial.

Figure 3 shows the responses to the same desired trajectory employing both the cascade controller and the proportional derivative with gains given by $P = 9 \times 10^{-3}$ and $D = 5 \times 10^{-3}$ (this gains correspond again to the limit of stability).

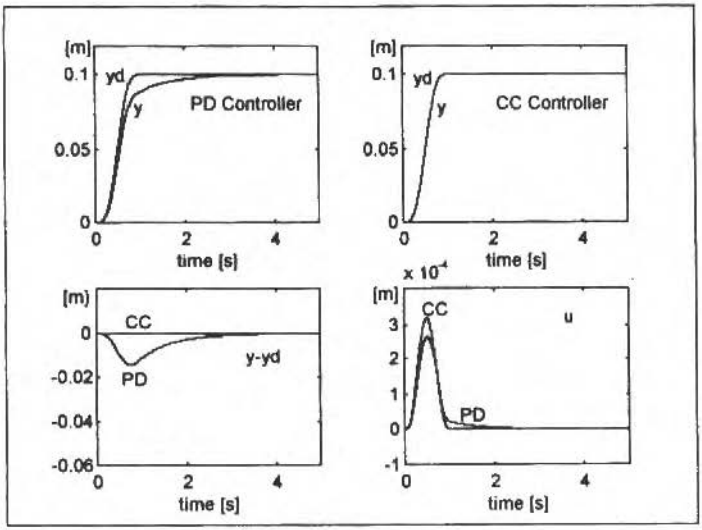


Fig. 3 Linear Model Responses to the PD and CC Controller with a Desired Trajectory Described by the 7th Order Polynomial.

For the regulation control problem these simulation results illustrate that the steady-state error ($y - y_d$) is zero with the P, PD and CC controllers. Using the PD controller we can reduce the maximum error with respect to the performance with the P controller. It can be observed that this error should remain considerable, because of the bandwidth limitations. We can also verify that the cascade controller leads to small errors (near zero) thanks to high gains.

In Fig. 4 we have the hydraulic actuator responses to the cascade and proportional controllers considering a desired trajectory given by $y_d(t) = 100 \sin(0.5 \pi t)$. For these simulation results we have $P = 4 \times 10^{-3}$.

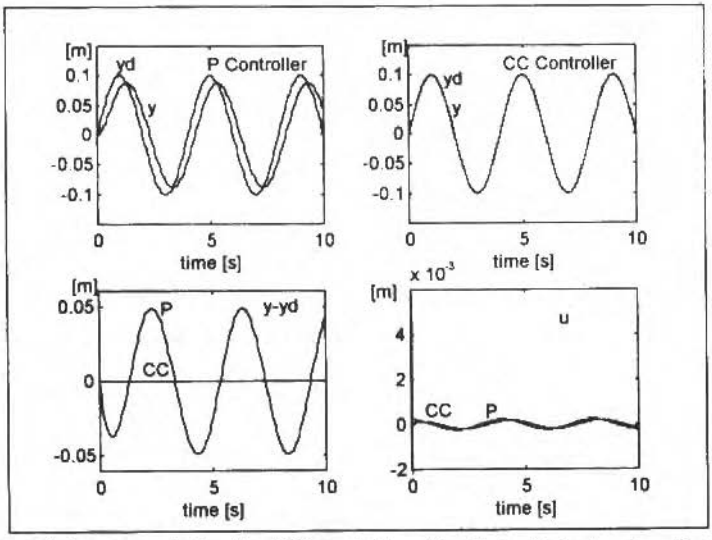


Fig. 4 Linear model responses to the P and CC controller with a desired trajectory described by $y_d(t) = 100 \sin(0.5 \pi t)$.

To the same trajectory, Fig. 5 presents the responses to the cascade and proportional derivative controller (with $P = 9 \times 10^{-3}$ and $D = 5 \times 10^{-3}$).

To the trajectory tracking problem Figs. 4 and 5 show that the P and PD controllers lead to a tracking error along the trajectory. This error is smaller with the PD controller, but it can be observed that it should remain considerable. This occurs again because of the bandwidth limitations. In the trajectory tracking situation we can also verify that the cascade controller propitiates small errors (near zero) thanks to high gains.

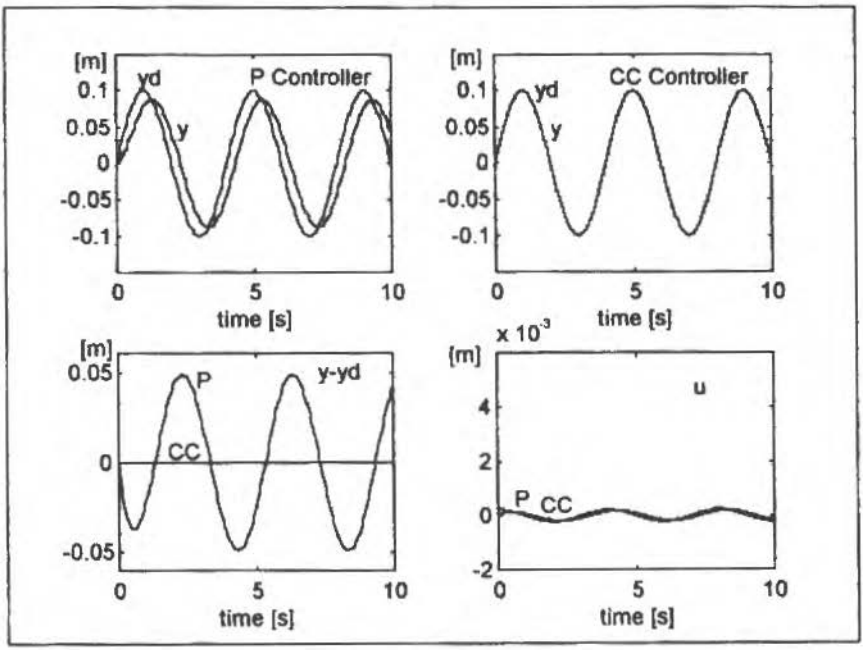


Fig. 5 Linear Model Responses to the PD and CC Controller With a Desired Trajectory Described by $y_d(t) = 100 \sin(0.5 \pi t)$.

Conclusions

In this paper, the problem of modeling and control of a hydraulic actuator has been considered. The main characteristics of the system have been analyzed and a control algorithm has been proposed in order to overcome detrimental effects on the dynamic responses of hydraulic actuators.

The advantage of the cascade control scheme is to improve the system performances when compared with classical controllers (e.g. PD and PI controllers). An important feature of this method is to ensure the exponential stability of closed loop error in the case of all parameters are known. The exponential stability of the closed loop system ensures robustness against external disturbances (like the load) and non modeled dynamics like the servovalve and the pipeline dynamics in the hydraulic actuator.

Simulation results of the cascade control scheme yield satisfactory behavior and good robustness properties.

Future research includes practical implementation of this method associated with more elaborate control design procedures: adaptive and robust control techniques applied to the mechanical and hydraulic components.

Acknowledgments

The authors wish to express his gratitude to CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico (Proc. No. 520944/95-0 (NV)).

References

Bodson, M., and Sastry, S., 1989, "Adaptive Control Stability, Convergence and Robustness", Englewood Cliffs, Prentice Hall.

Burton, T.D., 1994, "Introduction to Dynamic Systems Analysis", McGraw-Hill, Inc., New York, USA.

Davidson, E.J., 1990, "Benchmark Problems for Control System Design", IFAC Theory Committee Report, pp. 30.

Franklin, G.F., Powell, J.D., and Emani-Naeini, A., 1994, "Feedback Control of Dynamic Systems" 3th edition, Addison-Wesley Publishing Company, USA.

Guenther, R., and Hsu, L., 1993, "Variable Structure Adaptive Cascade Control of Rigid-Link Electrically-Driven Robot Manipulators", Proc. IEEE 32nd CDC, San Antonio, Texas, December, pp. 2137-2142.

Heintze, J., and Weiden, A.J.J., 1995, "Innerloop Design and Analysis for Hydraulic Actuators, With Application to Impedance Control", Control Engin. Practice, 3, pp. 1323-1330.

Hsu, L., and Guenther, R., 1993, "Variable Structure Adaptive Cascade Control of Multi-Link Robot Manipulators with Flexible Joints: the Case of Arbitrary Uncertain Flexibilities", Proc. IEEE Conf. Robot. Autom., Atlanta, Georgia, May, pp. 340-345.

Noble, B., and Daniel, J.W., 1977, "Applied Linear Algebra", Prentice Hall, Inc. Englewood Cliffs, New Jersey, USA.

Paim, C.C., 1997, "Técnicas de Controle Aplicadas a um Atuador Hidráulico", Master Thesis, Universidade Federal de Santa Catarina, in Portuguese.

Schothorst, G.V., Teerhuis, P.C., and Weiden, A.J.J., 1994, "Stability Analysis of an Hydraulic Servo-system including Transmission Line Effects", Proc. 3th Intern. Conf. Automation, Robotics and Computer Vision, Singapore, pp. 1919-1923.

Slotine, J.J.E., and Li, W., 1988, "Adaptive Manipulator Control: A Case Study", IEEE Trans. on Auto. Control, 33-11.

Slotine, J.J.E., and Li, W., 1991, "Applied Nonlinear Control", Prentice Hall Inc., USA.

Stringer, J., 1976, "Hydraulic Systems Analysis", The Macmillan Press, USA.

Vidyasagar, M., and Vannelli, A., 1982, "New Relationship between Input-Output and Lyapunov Stability", IEEE Trans. Auto. Control, vol. AC-27, no. 2, pp. 431-433.

Virtanen, A., 1993, "The Design of State Controlled Hydraulic Position Servo-system", Proc. of the 3rd Scandinavian Inter. Conf. on Fluid Power, pp. 193-206.

Utkin, V.I., 1987, "Discontinuous Control Systems: State of Art in Theory and Applications", Preprints IFAC 10th Word Congress on Automatic Control, Munich, vol. 1, pp. 75-94.

Watton, J., 1989, "Fluid Power Systems", Prentice Hall, Cambridge, England.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 121-137

ISSN 0100-7386
Printed in Brazil

Modeling and Simulation Results for a Natural Gas Internal Combustion Engine Coupled to a Hydraulic Dynamometer

Agenor de Toledo Fleury
José Augusto Lopes
Ney Ricardo Moscati
IPT - Instituto de Pesquisas Tecnológicas
Divisão de Mecânica e Eletricidade
Agrupamento de Sistemas de Controle, C.P. 7141
01064-970 São Paulo, SP Brasil

Francisco Emílio Baccaro Nigro
Maurício Assumpção Trielli
Agrupamento de Motores C.P. 7141
01064-970 São Paulo, SP Brasil

Abstract

When considering the problem of pollutant emission reduction from internal combustion engines, many technical challenges arise. One of them is the control strategy formulation to be applied to the fuel injection and ignition systems, with the aim at regulating the air-fuel ratio around the stoichiometric value. One of the steps towards this goal is the development of an engine model. This model has to be validated in a testbed, which includes a dynamometer and a gas analyser. IPT Engines Lab is equipped with a hydraulic dynamometer, whose dynamics is slower than the engine one. Modeling and simulation results for these two coupled models antecipate results for testbed operation and helps to understand and control the engine transient, where the dynamometer dynamics is to be excluded from the achieved experimental results.
Keywords: *Internal Combustion Engine, Hydraulic Dynamometer, Mathematical Modeling, Engine Testbed Simulation.*

Introduction

The appearance of stringent laws related to pollutant emissions from internal combustion engines (ICE) has been the driving factor for the development of modern engines and for the investigation on the use of alternative fuels. Besides that, many control technologies have been added to fuel injection and ignition systems in order to achieve the desired figures of pollutant emission reduction (Cassidy *et al.*, 1980; Aquino, 1981; Jones *et al.*, 1988; Noble and Beaumont, 1991; Moskwa, 1993; Dan Cho and Oh, 1993; Abida and Claude, 1994; Chang *et al.*, 1995). Although this problem seems to be solved for steady-state engine operation, and even there are some off-the-shelf solutions for small car engines (Kaiser *et al.*, 1988), the lack of feasible solutions for transient engine dynamics (for medium or large size engines) has made it a quite attractive topic of research nowadays (Fleury and Lopes, 1994). It is important to remember that the São Paulo City traffic legislation points out that in about 10 years all bus fleet (around 12,000) must run on natural gas engines in order to reduce downtown pollution levels.

This work is part of a large research project under development at IPT with the aim at developing results to make the use of natural gas on urban buses engines feasible. This project comprises four sub-items: a thermodinamic simulator for engine design, the design of new combustion chambers for natural gas use, studies on the influence of natural gas composition over engine performance and studies on natural gas fuel injection and ignition control systems performance under transient conditions.

Given this context, this work presents the development of a dynamical model for a spark-ignited internal combustion engine coupled to a hydraulic dynamometer. Dynamometers have been used for decades to analyse and calibrate ICEs under specified steady-state conditions. When one has to deal with transient operation, the engine-dynamometer dynamics influences torque and rotation measurements. Then, for identification of basic transient engine parameters, it is of fundamental importance to distinguish between engine and dynamometer contributions to the complete system

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editor: Hans Ingo Weber.

dynamical response. Two case studies are considered. First, the engine equipped with an ideal gas mixer is simulated; second, a fuel injection/ignition system controlled through a predictive scheme (GPC Controller) substitutes for the mixer, looking for a near future practical application.

Problem Statement

Natural gas is one of the most important clean alternative fuels nowadays, because it has no SO_2 and particulate emission and exhibits low emission rates for CO and HC. Unfortunately, the NO_x emission rate is still high and must be controlled in some way.

The development of natural gas engines to meet the stricter than ever emissions requirements can use two distinct approaches: use of stoichiometric mixture with 3-way catalyst; or use of lean-burn mixture with turbocharger and aftercooler. In most cases, ICEs are being developed to operate with stoichiometric mixture, where the reduction or oxidation of emission gases is performed by a 3-way catalitic converter (Fig.1). This means a serious control problem, because a 1% deviation in the air-fuel ratio (relative to the stoichiometric rate) may correspond to a 50% degradation in the converter efficiency (Dan Cho and Oh, 1993). In this way, there is need for a very precise fuel injection/ignition control system, to be implemented by electronic (microprocessed) means. A complete scheme for the solution of this problem is shown in Fig. 2.

Our current challenge (not yet fully implemented) is the design of a natural gas spark-ignited engine simulator, from a system dynamics point of view, in order to allow plant, actuators and sensors simulation and control strategies development, specially in transient conditions. This simulator is being developed while IPT's new Engines Lab is in implementation phase, with an experimental set-up consisting of a gas converted engine, a hydraulic dynamometer and a gas analyser (Fig. 3).

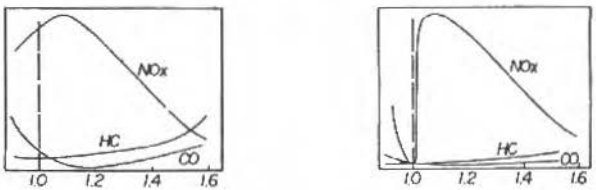


Fig. 1 Typical Emissions (a) Without and (b) With Catalitic Converter

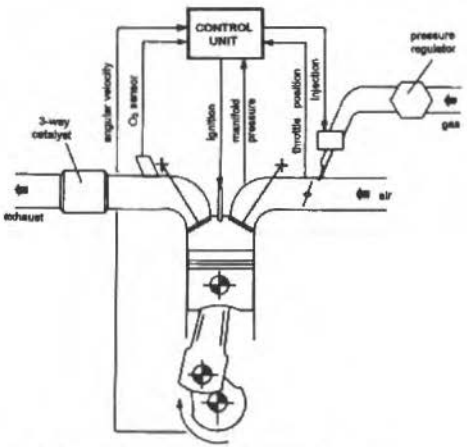


Fig. 2 Typical Scheme for Engine Control

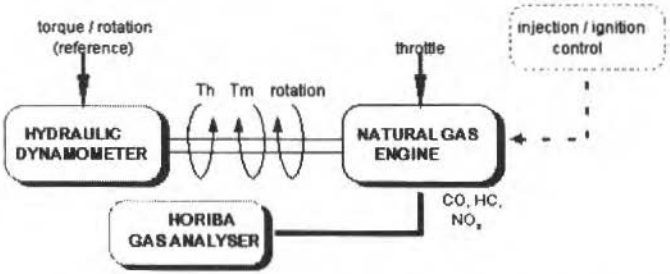


Fig. 3 Experimental Set-up Scheme

Keeping in mind the importance of the transient analysis and the absence of off-the-shelf controllers for use on natural gas bus engines, performance analysis of some candidate control strategies assumes fundamental role. In this way, there is need for good engine models. On the other side, engine model validation has to be carried out at engines laboratories before implementation in on-board computers. Results concerning dynamometer modeling are very scarse in literature: Powell (1978) has presented an engine model considering the dynamometer as a variable load, and Isermann *et al.* (1993) has developed a complete identification model of an electrical dynamometer. Since there is need also for a good dynamometer model and no hydraulic dynamometer model is available in the literature, a decision was taken to create a simulator for engine-dynamometer set consisting of a non-linear simplified engine model, a non-linear parametric dynamometer model and fuel injection/ignition actuators and sensors models (under construction), in order to understand the complexity of the problem. In the sequence, engine-dynamometer set modeling and simulation shall be presented pointing out that for consistent transient engine results dynamometer responses must be considered. The simulator, whose parameters must be identified through specific tests in near future shall allow the development of advanced control techniques, as GPC (Clarke *et al.*, 1987), and the forecast of new components performance before laboratory tests, thus saving time and financial resources.

Engine Model

The simplified engine simulation model for control and identification purposes has been designed under the general scheme shown in Fig. 4. A modular representation has been used to describe its main subsystems: intake manifold, combustion and rotational dynamics, with typical parameter values assumed when empirical data have not yet been identified.

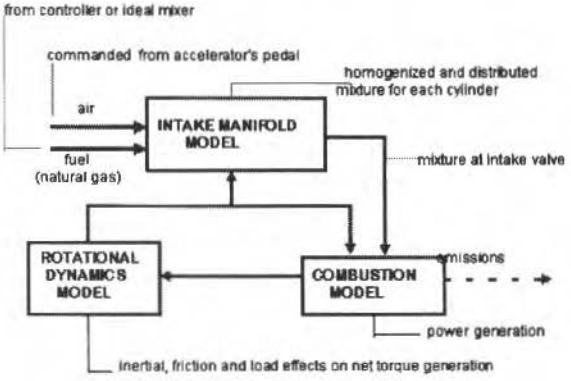


Fig. 4 Engine Simulator Scheme

Intake Manifold Model

The intake manifold model is described by its air and fuel dynamics. The main hypothesis are unidimensional and isoentropic compressible air flow (ideal gas) through the throttle, constant discharge coefficient at the throttle and uniform pressure and temperature distribution at the manifold.

The throttle area is modeled as shown in Eq. (1).

$$(\alpha)=\frac{\pi D^2}{16}(3-2\cos\alpha-\cos^2\alpha) \tag{1}$$

where α is the throttle angular position, $A(\alpha)$ is the cross-sectional area of the flow and D is the throttle diameter. Once calculated $A(\alpha)$, the air mass flow through the throttle can be obtained under two possible flow conditions (Moskwa and Hedrick, 1992):

i) If $\frac{P}{P_0}>\left(\frac{2}{k+1}\right)^{\frac{k}{k-1}}$,

$$\dot{m}_{ar_b}=C_d * A(\alpha)*\frac{P_0}{\sqrt{R*T_0}}*\left(\frac{P}{P_0}\right)^{\frac{1}{k}}*\left(\frac{2}{k+1}\right)^{\frac{k+1}{k-1}}*\left\{\frac{2k}{k-1}*\left[1-\left(\frac{P}{P_0}\right)^{\frac{k-1}{k}}\right]\right\}^{\frac{1}{2}} \tag{2}$$

ii) If $\frac{P}{P_0}\le\left(\frac{2}{k+1}\right)^{\frac{k}{k-1}}$ (blocked flow),

$$\dot{m}_{ar_b}=C_d * A(\alpha)*\frac{P_0}{\sqrt{R*T_0}}*k^{\frac{1}{2}}*\left(\frac{2}{k+1}\right)^{\frac{k+1}{2(k-1)}} \tag{3}$$

where $\dot{m}_{ar_b}$ is the air mass flow, P_0 e T_0 are the stagnation pressure and temperature respectively, P is the intake manifold pressure, C_d is the discharge coefficient through the throttle, R and k are gas thermodynamic constants for the air.

The air dynamics inside the manifold is based on the continuity law , defined by the air mass flow inlet $\dot{m}_{ar_b}$ and the air mass flow to the cylinders, $\dot{m}_{ar_v}$ (Eq. 4).

$$m_{ar_v}=\frac{N_{cil}*\omega}{4\pi}*\frac{P*V_{cil}}{R*T}*\eta_v \tag{4}$$

Here N_{cil} is the number of cylinders, V_{cil} is the cylinder volume, n is the engine angular velocity, P is the intake manifold pressure, T is the intake manifold temperature and η_v is the engine volumetric efficiency. Then, the air pressure dynamics can be written as in Eq. (5).

$$\dot{P} = (\dot{m}_{ar_b} - \dot{m}_{ar_v}) * \frac{R * T}{V} \tag{5}$$

where V is the intake manifold volume.

The fuel dynamics that define the fuel flow to the cylinders corresponds to a diffusion process on the intake manifold. A simpler representation is normally used to keep the system integrable. This is done by adding just a smoothing block and a transport delay block (Dobner, 1983) to the fuel injected at the nozzle. The fuel dynamics model is then simplified to the Eq. (6).

$$\dot{m}_f = \frac{1}{1 + T_{sm}s} * e^{-sT_f} * \dot{m}_{f_{inj}} \tag{6}$$

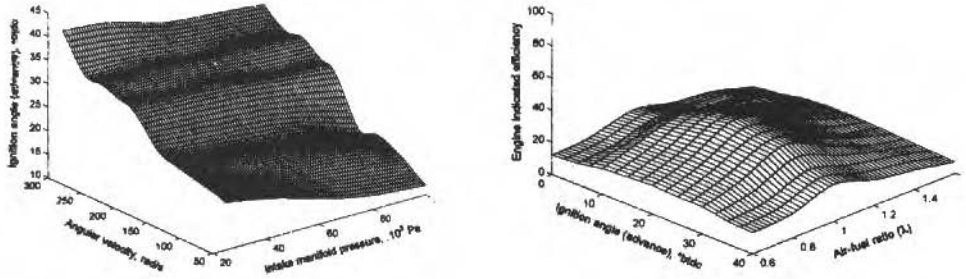
where T_{sm} is the smoothing time constant (assumed 0.01 s), T_f is the fuel transport delay (assumed 0.01 s) and $\dot{m}_{f_{inj}}$ is the fuel mass flow defined by the ideal gas mixer or by the GPC controlled injection system (Lopes, 1996).

Combustion Model

Chemical and thermal processes ocurring in the combustion chamber for the transformation of the fuel chemical energy to torque and/or angular shaft velocity are highly complex and the corresponding chemical and thermal models are not useful for dynamic simulation and control. From a system dynamics point of view there are two important phenomena to consider. The first one corresponds to the torque generation and the second corresponds to the exhaust gas transportation from the combustion chamber to the catalitic converter inlet where the oxygen sensor (λ sensor) measures the air-fuel ratio. Then, the combustion model comprises two blocks that allow to calculate the indicated torque, based on ICE efficiency and ignition angle maps, and to include a variable transport delay for the air-fuel ratio, corresponding to the ICE 4-stroke cycle. ICE torque is modeled as:

$$T_{ind} = \frac{\dot{m}_f * H * \eta_{ind}}{\omega} \tag{7}$$

where T_{ind} is the indicated torque (Nm), H is the lower heating value for natural gas (assumed 4.10^7 J/kg), ω is the engine angular velocity (rad/s) and η_{ind} is the indicated engine efficiency, function of the air-fuel ratio and of the ignition angle. Typical surfaces assumed for η_{ind} and ignition angle as empirical static maps are shown in Fig. 5.



(a) (b)

Fig. 5 Typical Ignition Angle (a) and Indicated Engine Efficiency Surfaces (b)

The air-fuel ratio measured at the λ-sensor is approximated by:

$$\lambda_{ls} = \lambda_{im} * e^{T_v s} \tag{8}$$

where λ_{ls} is the air-fuel ratio from intake manifold dynamics, λ_{im} is the air-fuel ratio to lambda sensor and T_v is a variable time constant proportional to the inverse of the angular speed.

Rotational Dynamics

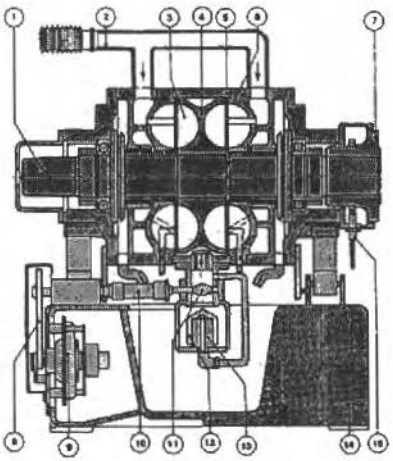
The rotational dynamics module is used to calculate shaft angular acceleration, $\dot{\omega}$, due to changes in the preceding modules. It includes the engine and dynamometer inertias and typical friction and load components when producing angular acceleration from net torque, that represents the difference between indicated and hydraulic torques (Eq. 9).

$$T_{ind} - \Im - L = J_t * \dot{\omega} \ . \tag{9}$$

where $\Im$ is the friction torque, calculated as a linear function of the angular velocity n, L is the hydraulic torque from dynamometer model, J_t is the total inertia (engine + dynamometer, kg.m²) and ω is the angular velocity, that is used in combustion and intake manifold models.

Dynamometer Model

IPT engines laboratory is equipped with a hydraulic dynamometer Schenck model D360-1e, which is usually used to evaluate the engines performance under steady state conditions. A schematic dynamometer view is shown in Fig.6.



1 - Rotor shaft;
2 - Water inlet;
3 - Whirl chambers;
4 - Double rotor;
5 - Housing;
6 - Stator;
7 - Coupling flange;
8 - Toothed belt drive;
9 - Drive motor for control valve;
10 - Coupling shaft;
11 - Control valve throttle;
12 - Piston valve;
13 - Guideway;
14 - Dynamometer basis;
15 - Speed transducer;

Fig. 6 Schematic View of the Hydraulic Dynamometer SCHENCK D360-1e (Schenck 1)

The dynamometer operates based on the Föttinger principle. Water from a supply tank [2] flows to the whirl chambers [3] in a way that brakes the rotor and cools the system. The heated water flows off through the control valve [11,12]. The resistant torque that the dynamometer offers to the engine is a function of the water level, which is controlled by a valve, among other variables. The housing, supported via a lever on the force measuring equipment, rotates thus indicating the corresponding torque.

The dynamometer can operate in two basic modes: rotation or torque tracking. In the first mode, feedback is provided by an angular velocity transducer, while in the second one, feedback is obtained through a load cell. In both cases a PID controller closes the loop.

The hydraulic dynamometer is a highly complex system which involves complicated processes as flow turbulence at the whirl chambers, forces actuating on the piston valve, housing movement, etc. Based on some simplifying assumptions as disregarding the local effects of the turbulent flows and based on fundamental principles, a parametric mathematical model has been built (Moscati *et al.*, 1996). This model still remains a nonlinear complex one as shown in the sequence.

As water circulates in the chambers, it receives momentum at the rotor side of them and looses it in the stator side of the chambers, what causes a braking torque on the rotor. Simultaneously, kinetic energy is transfered to the water; is then transformed into thermal energy through turbulence, and heats the water.

Assuming a water layer circulating in the periphery of the chambers, with a thickness ΔR over the most external radius R of the chambers, one can estimate the pressure caused by the water layer as:

$$\Delta P = \rho.\omega.^2 R_m.\Delta R \tag{10}$$

where R_m is the mean radius of the water layer $\left(R_m = R - \frac{\Delta R}{2} \right)$ and ω is the angular velocity of the dynamometer rotor.

This same pressure is assumed to promote the flow off the dynamometer through the control valve.

By applying the Continuity Law to the chamber, one can obtain the time variation of the water film height inside the dynamometer as:

$$\Delta Q = \frac{d}{dt}(2\pi.\ell.R_m.\Delta R) = Q_e - Q_s \tag{11}$$

where ΔQ is the water volume rate within the chambers, Q_e and Q_s represent the water inlet and outlet volume flow rates and ℓ is the chamber width.

The hydraulic torque T_C transmited to the housing by the water can be estimated by applying Euler's moment of momentum equations to the fluid, while using values of head loss factors based on hydraulic pumps and turbines:

$$T_C = \rho.R^3.\omega^2.2\pi.R_m.\Delta R \tag{12}$$

This hydraulic torque reacts in order to brake the rotor, and one can use the Moment of Momentum Theorem to write:

$$T_m = J_d.\dot{\omega} + T_C \tag{13}$$

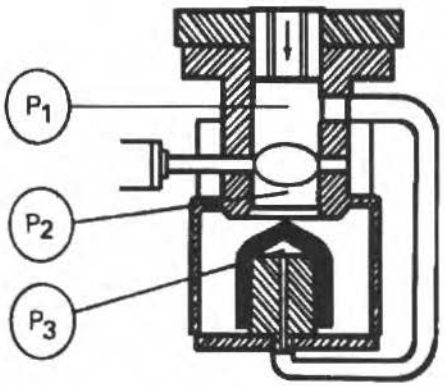
where $\dot{\omega}$ is the angular acceleration of the rotor, T_m is the efective engine torque $(T_{ind} - \mathfrak{I})$ and J_d is the moment of inertia of the dynamometer shaft and rotor.

Assuming that the housing resistant torque is equal to hydraulic torque T_C and considering the housing-arm set as a second order system (inertia-torsional spring-damper), the housing-arm movement can be described by:

$$T_C = J_c.\ddot{\theta}_c + C_T.\dot{\theta}_c + K_T.\theta_c \qquad (14)$$

where J_c is the housing-arm set moment of inertia, C_t is an equivalent damping constant for the set, K_T is an equivalent spring constant, θ_c is the housing angular displacement relative to the static equilibrium and θ_c e θ_c are its angular speed and acceleration, respectively. Note that, as the housing is supported via a lever on the measuring equipment, the load cell signal is proportional to the housing angular displacement.

The above equations are dependent on the volume of water inside the dynamometer. As the water inlet flow is constant, the internal volume is regulated by a complex control valve at the dynamometer outlet. It is composed by a butterfly valve, actuated through a DC motor that is part of the dynamometer control loop, and a hydraulic pressure attenuation device. This device consists of a hydraulic circuit that transmits pressure from the chamber just before the throttle to the internal side of a piston. The piston is free to move between two stops, acording to the differential pressure between its internal chamber and the chamber before the throttle. The piston movement regulates the area for water flow outside the dynamometer. The control valve and its different pressure chambers are shown in Fig. 7.



p_1 - Dynamic pressure before throttle;

p_2 - Pressure after throttle;

p_3 - Pressure in piston internal chamber

Fig. 7 Outlet Flow Control Valve

The flow through the control valve, corresponding to the outlet flow from the dynamometer, Qs, is given by:

$$Qs = \frac{C_d A(x)}{\sqrt{1-\left(\frac{4C_d}{\pi D^2 A(x)}\right)^2}} \sqrt{\frac{2p_2}{\rho}} \tag{15}$$

where C_d is the discharge coeficient assumed constant ($C_d = 0.67$) and A(x) is the water flow area as a function of the piston position x:

$$A(x) = \begin{cases} \pi D_s^2 & 0 \le x \le 10.5e-3 \quad (m) \\ \pi D_s^2 - \frac{D_s^2}{4}\left[sin\theta + 2\pi - \theta\right] & 10.5e-3 \le x \le 27.5e-3 \quad (m) \\ 0 & x \ge 27.5e-3 \quad (m) \end{cases} \tag{16}$$

In Eq. (16),

$$\theta = 2\,arccos\left(\frac{x-0.019}{0.0085}\right) \tag{17}$$

and D_s is the diameter of the four outlet holes uncovered by the piston.

Assuming that the outlet flow can be approximated by the flow through the throtlle, thus disregarding the flow necessary to move the piston, and using Eq. (15), one can write the pressure p_2 as:

$$p_2 = \frac{C_v^2 A_v^2}{F(x) + C_v^2 A_v^2} p_1 \tag{18}$$

where:

$$F(x) = \frac{C_d^2 A(x)^2}{F_d(x)} \tag{19}$$

$$F_d(x) = 1 - \left(\frac{4C_d}{\pi D^2} A(x)\right)^2 \tag{20}$$

$$C_v A_v = \left[\frac{0.78(1-\cos\alpha)}{1-0.78(1-\cos\alpha)}\right]\frac{\pi D^2}{4} \tag{21}$$

and α is the butterfly angular position and D is the piston diameter.

The pressure p_3 is calculated considering the pressure drop along the tube that links the chamber before the butterfly valve to the internal piston side:

$$p_3 = p_1 - 9.06*10^5 * \dot{x}|\dot{x}| \tag{22}$$

Note that, as water can flow in both directions according to piston travel direction, a module function has to be included in Eq. (22) to describe situations where pressure p_1 is greater than p_3.

In order to obtain the piston position x, one has to solve the following force balance equation:

$$F_{total} = m\ddot{x} = p_3 A_{ef} - mg - F_{perm} \tag{23}$$

where A_{ef} is the efective area under the piston, mg is the piston weight and F_{perm} is the steady state component of the reaction force to the change in flow momentum, due to variation in flow direction.

The water level inside the whirl chamber is controlled through an usual PID controller (ISA algorithm), that commands the butterfly valve angle, $\alpha(t)$, as a function of the error signal e(t):

$$\alpha(t) = K_p[e(t) + T_d \frac{de(t)}{dt} + \frac{1}{T_i}\int e(t)dt] \tag{24}$$

where the parameters K_p (proportional gain), T_d (derivative time) and T_i (integral time) can be tuned in the dynamometer control module.

When the dynamometer is operating in the rotation tracking mode, the error signal e(t) is the difference between a rotation reference signal and the rotation measured by an angular velocity transducer. In the torque tracking mode, the error signal is the difference between a preseted torque and the torque measured by the load cell. In both cases, the control signal commands the control valve, increasing or decreasing the water level inside the whirl chamber.

Complete Simulation Scheme

Since engine and hydraulic dynamometer are coupled through a rigid shaft, the resulting torque produces changes in angular shaft velocity that are feedback for both models. In terms of external inputs, the engine torque depends basically on engine throttle position time behaviour and dynamometer torque depends on the desired reference for torque or rotation. In this way, different operating points can be achieved as defined previously. Finally, the initial conditions for manifold pressure (in the engine model), water film level at whirl chamber, control valve angular position , piston position and velocity (in the dynamometer model), and shaft angular velocity are necessary for simulation. The complete simulation scheme is illustrated in Fig. 8.

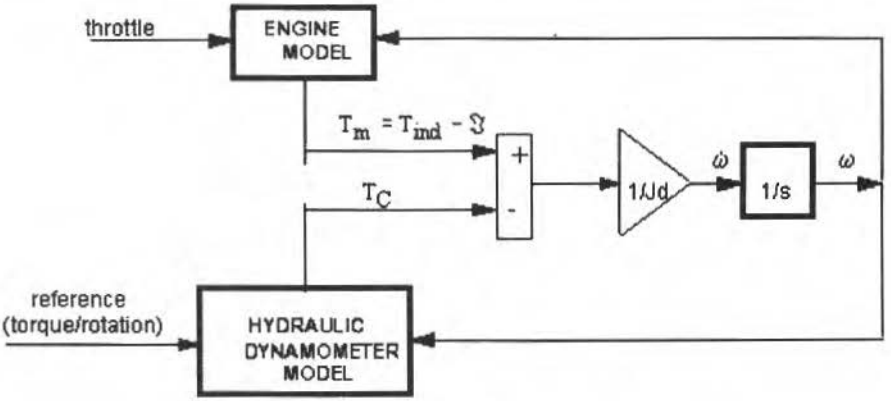


Fig. 8 The Testbed Simulation Scheme

Simulation Results

Two fuel injection conditions have been simulated. The first one considers an ideal gas mixer, i.e., a mixer that instantaneously injects the necessary amount of gas to maintain a stoichiometric mixture ratio. In the second case , the injected fuel mass flow and ignition angle have been determined using a predictive control strategy given by a multivariable GPC (Generalized Predictive Controller). A detailed description of this controller and its tunning to this particular problem are shown in the work of Lopes (1996).

The ideal gas mixer represents the simplest condition for this problem, although not feasible for practical implementation. The GPC controller is seen as a candidate for future applications on natural gas ICE fuel injection and ignition control systems.

Fig. 9 shows the results related to an ICE equipped with an ideal gas mixer system where the aim is to track a torque reference of 325 Nm up to 10 s, a linear increase of 100 Nm between 10 and 20 s, and a 425 Nm reference up to 40 s. This is the set-point profile sent to the dynamometer PID controller operating on the torque tracking mode. A transient engine acceleration is applied to the engine throttle by changing it from 68° to 80° at t=28 s.

The first 4 seconds are used to reach steady-state operation as can be seen in Figs. 9 (d)-(j). In the range of the linear increasing torque reference (10 to 20 s), engine power requirement is fixed because there is no throttle command. Then, the shaft angular velocity decreases, while manifold pressure is increasing since it is a measure of engine load profile. On the other side, the hydraulic torque reference is associated with water resistance imposed to the rotors and, as consequence, the brake water level increases as butterfly and piston valves close.

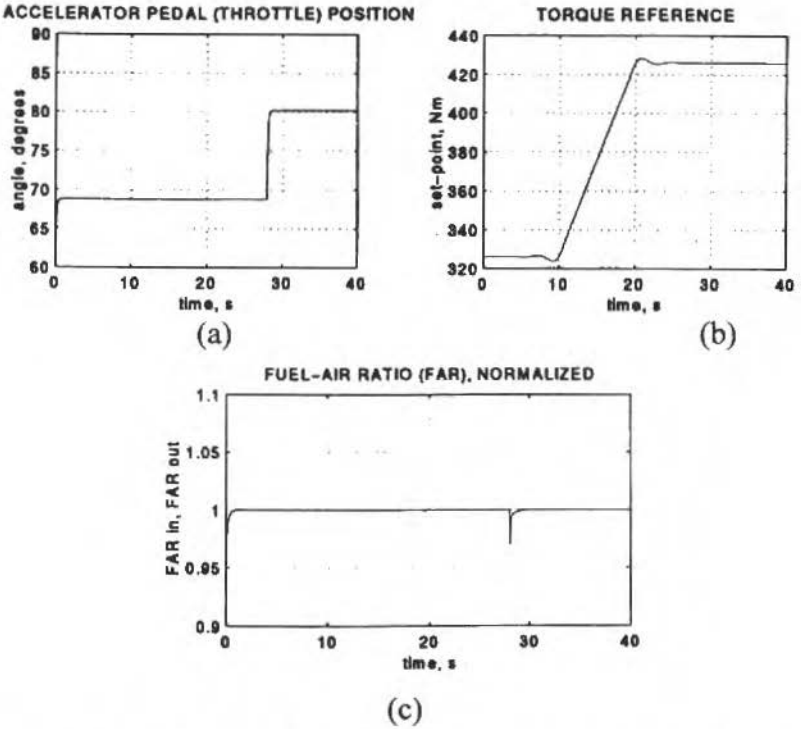
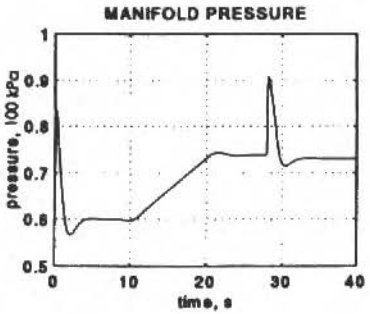
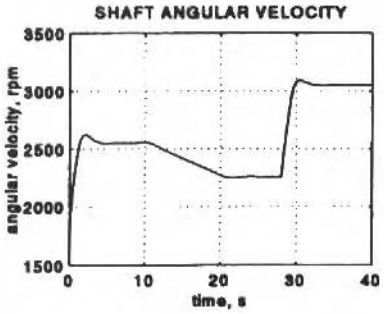


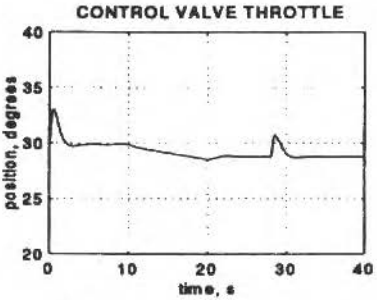
Fig. 9 Simulation #1: Ideal Gas Mixer With 10° Throttle Transient Inputs: (a) and (b); Engine Variables: (c), (d), and (e); Dynamometer Variables: (f), (g), (h), (i); Torques: (j) and (k). (cont. on next page)



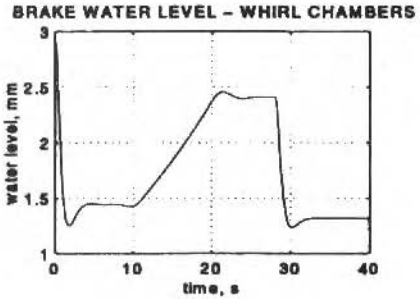
(d)



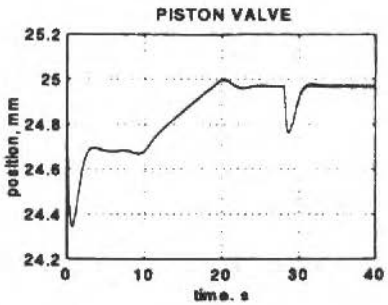
(e)



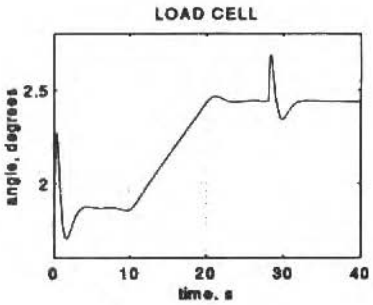
(f)



(g)

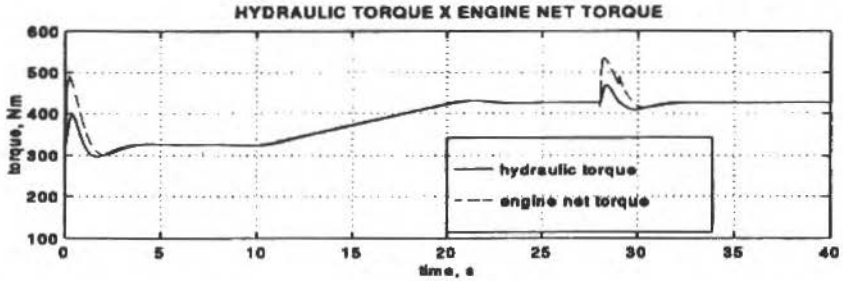


(h)

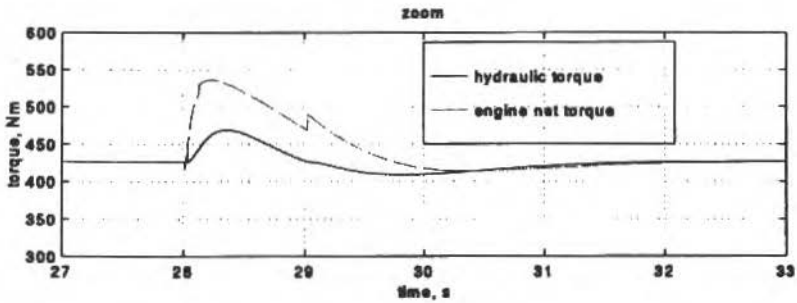


(i)

Fig. 9 Simulation #1: Ideal Gas Mixer With 10° Throttle Transient Inputs: (a) and (b); Engine Variables: (c), (d), and (e); Dynamometer Variables: (f), (g), (h), (i); Torques: (j) and (k). (cont. on next page)



(j)



(k)

Fig. 9 Simulation #1: Ideal Gas Mixer With 10° Throttle Transient Inputs: (a) and (b); Engine Variables: (c), (d), and (e); Dynamometer Variables: (f), (g), (h), (i); Torques: (j) and (k).

At t=28 s, the engine throttle transient causes an increase in engine power and a small deviation from the stoichiometric fuel-air ratio at the λ-sensor as expected by the engine model, which considers a delay in the fuel path, Fig. 9 (c). The PID reacts to maintain the hydraulic torque at set-point, thus increasing the shaft angular velocity and trying to hold the manifold pressure constant. This is done by opening the control valve for a short period of time and then decreasing the whirl chamber water level. This is the only way to maintain constant hydraulic torque when there is an increase in the shaft angular velocity. Note that the piston operates as a servo, tracking the butterfly valve movements of closing and opening water outlet.

Figures 9 (j) and (k) demonstrate that under steady-state conditions the torques are equal and dynamometer gives a correct measure for the engine torque. This is the normal dynamometer operating condition. Nevertheless, in transient operation the dynamometer torque information is smoother than the engine one, as expected, and as one can measure only dynamometer response the achieved torque profile is fundamental to analyse and understand engine transient behaviour. Just in case of available good engine-dynamometer models, engine transient parameters can be derived from measured load cell dynamometer responses.

In Figure 10, a GPC controlled injection system is simulated substituting for the ideal gas mixer. Test profile consists of linearly increasing angular velocity reference, from 2100 rpm to 2400 rpm, during 20 s. A 10° engine throttle transient condition is imposed at t=10 s.

A linear increase in the angular velocity reference implies in manifold pressure and hydraulic torque decrease, because power requirement is constant (no throttle transient). In this case, the brake water level in the whirl chambers decrease, commanded by the control valve opening. When the throttle transient is applied, the opposite happens. Because of power instantaneous increase, there is

an almost instantaneous manifold pressure and engine torque increase. In consequence, angular velocity and hydraulic torque also increase. To maintain the shaft angular velocity close to the reference, the PID acts in order to reduce angular velocity through control valve closing, thus increasing the brake water level. After this period, all the variables continue to follow dynamometer profile. On the engine side, the GPC controller stabilizes the fuel-air ratio around 0.4 s after the pedal command, with a maximum fuel-air ratio excursion of around 5% of the stoichiometric value. In the sequence, all returns to the later condition.

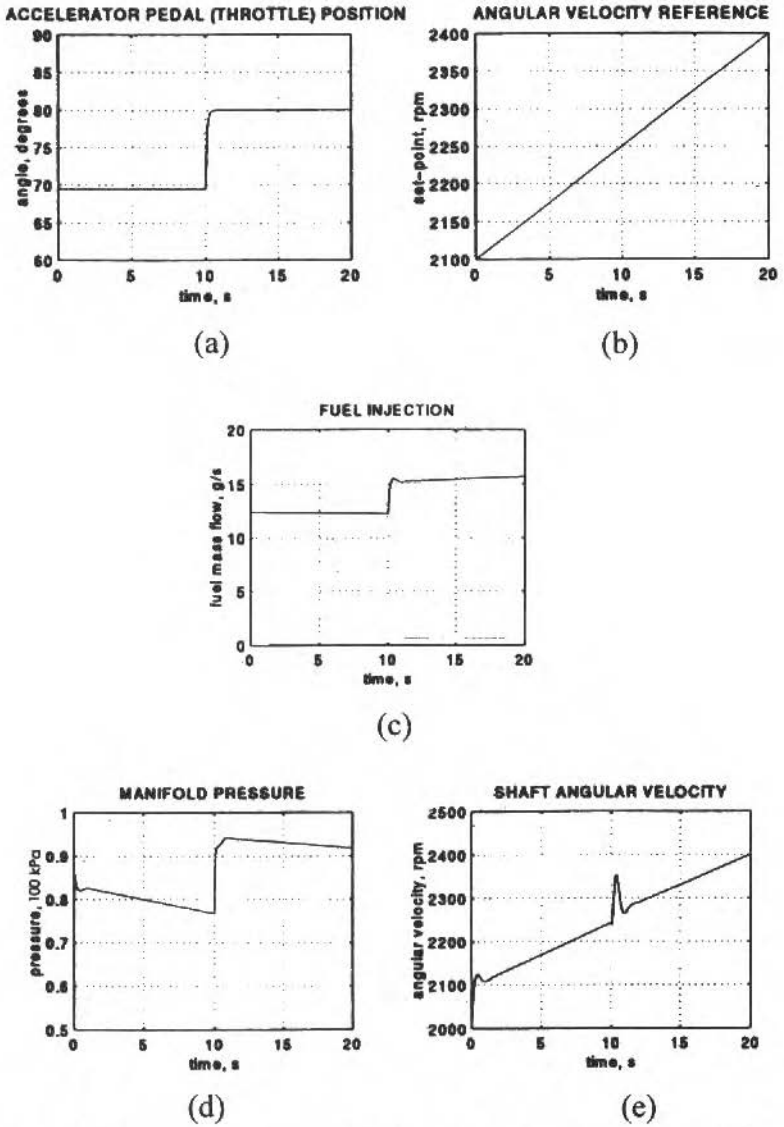
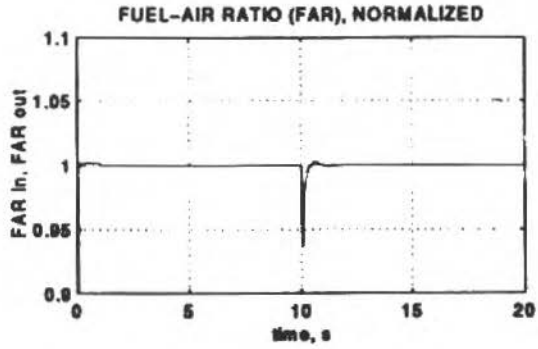
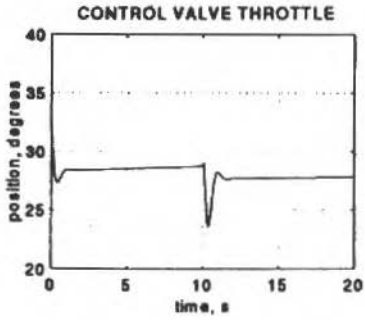


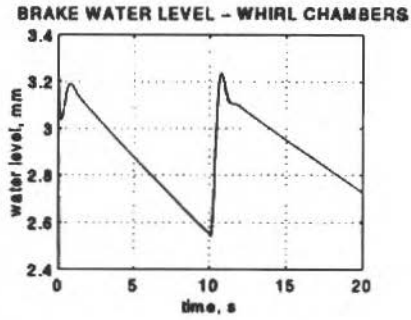
Fig. 10 Simulation #2: GPC Controlled Gas Injection With 10° Throttle Transient. Inputs: (a) and (b); Engine Variables: (c), (d), (e) and (f); Dynamometer Variables: (g), (h), (i) and (j); Torques: (k) and (l).
(cont. on next page)



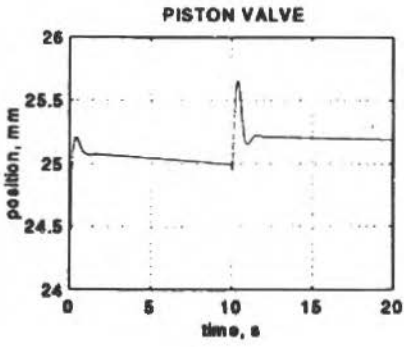
(f)



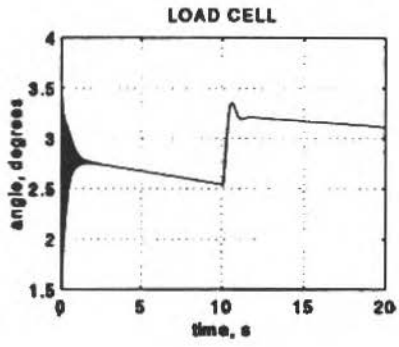
(g)



(h)



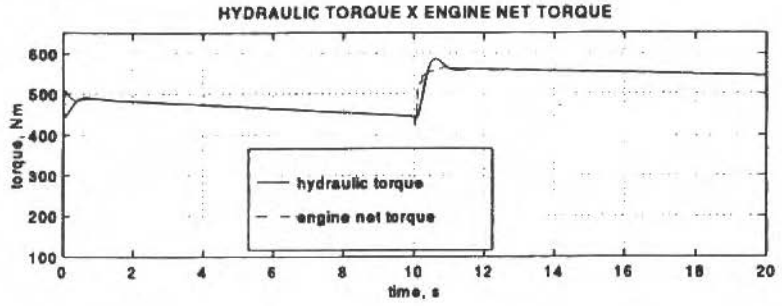
(i)



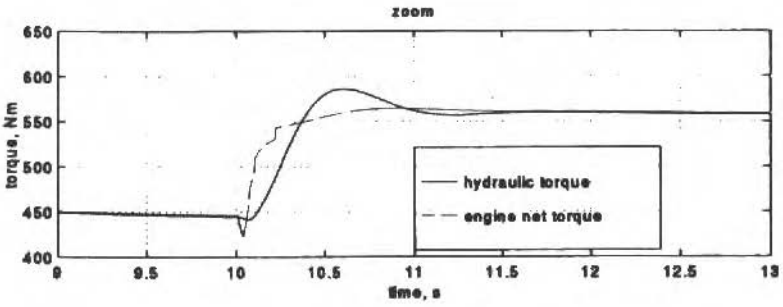
(j)

Fig. 10 Simulation #2: GPC Controlled Gas Injection With 10° Throttle Transient. Inputs: (a) and (b); Engine Variables: (c), (d), (e) and (f); Dynamometer Variables: (g), (h), (i) and (j); Torques: (k) and (l).

(cont. on next page)



(k)



(l)

Fig. 10 Simulation #2: GPC Controlled Gas Injection With 10° Throttle Transient. Inputs: (a) and (b); Engine Variables: (c), (d), (e) and (f); Dynamometer Variables: (g), (h), (i) and (j); Torques: (k) and (l).

It can be seen that, during transient, the PID action allows a 100 rpm deviation from the reference. Furthermore, and very important, the deviation from the hydraulic torque measurement to the real engine net torque, at t=10.1 s, is around 75 Nm. This is crucial for the dynamic prediction of the engine environment.

Conclusions

In this work, the importance of the gas pollutant emissions control from ICEs has been described. Although this topic has made a lot of progress in the last decade, much rests to be done and/or implemented in practical cases. Available injection control systems provide emission reduction based only on steady-state conditions. A more stringent emission control is depending on engine transient analysis through very expensive tests, demanding expensive new equipments. Our effort points towards the improvement of available equipments, modeling and identification of their dynamical interactions, in order to understand transient engine operation, and design of controllers capable of efficient emission reduction.

A natural gas ICE and a hydraulic dynamometer dynamic models have been presented. Two simulation have been shown, the first one considering an ideal gas mixer and torque reference tracking, and the second considering a GPC fuel controlled injection system and engine shaft angular velocity tracking. Torque prediction errors during transient have exceeded 50 Nm. This justifies deeper studies for a serious engine control research program. Future developments, like the study of other control strategies for fuel injection and ignition systems or for dynamometer control (substituting PID to get transient performance improvement), and even complete control

(simultaneous torque and rotation tracking) shall be necessary. The next steps are the validation of the engine and dynamometer models including sensors and actuators models in laboratory.

Acknowledgements

The authors thank FAPESP- Fundação de Amparo à Pesquisa do Estado de São Paulo - for financial support of part of this Project under Grant 93/0566-9 and CNPq- Conselho Nacional de Desenvolvimento Científico e Tecnológico - for financial support of José Augusto Lopes through a RHAE/MCT Program Scholarship.

References

Abida, J., and Claude, D., 1994, "Spark Ignition Engines and Pollution Emission: New Approaches in Modelling and Control." Int. J. of Vehicle Design, vol.15, nº 3, pp. 494-508.

Aquino, C.F., 1981, "Transient A/F Control Characteristics of the 5 Liter Central Fuel Injection Engine". SAE paper 810494.

Cassidy, J.F., Athans, M., and Lee, W.-H., 1980, "On the Design of Electronic Automotive Engine Controls Using Linear Quadratic Control Theory". IEEE Transactions on Automatic Control, vol. AC-25, nº 5, pp.901-912.

Chang, C.-F., Fekete, N.P., Amstutz, A., and Powell, J.D., 1995, "Air-fuel Ratio Control in Spark-Ignition Engines Using Estimation Theory". IEEE Transactions on Control Systems Technology, vol. 3, nº 1, pp.22-31.

Clarke, D.W., Mohtadi, C., and Tuffs, P.S., 1987, "Generalized Predictive Control - part I: The Basic Algorithm and part II": Extensions and Interpretations, Automatica, vol. 23, nº 2, p.137-148 and pp.149-160.

Dan Cho, D.-IL, and Oh, H.-K., 1993, "Variable Structure Control Method for Fuel-Injected Systems", Transactions of the ASME: Journal of Dynamic Systems, Measurement and Control, vol.115, pp.687-693.

Dobner, D.J., 1983, "Dynamic Engine Models for Control Development - Part I: Non-Linear and Linear Model Formulation", Int. J. of Vehicle Design, Technological Advances in Vehicle Design Series, SP4, Application of Control Theory in the Automotive Industry, pp. 54-74.

Fleury, A.T., and Lopes, J.A., 1994, "A Review on the Control Approaches to the Problem of Reducing Emissions From Natural Gas Internal Combustion Engines" (in Portuguese) In: Congresso Brasileiro de Automática, Rio de Janeiro, RJ, Anais. pp.488-493.

Isermann, R., Voigt, K.U., and Pfeiffer, K., 1993, "Parameter Estimation and Digital Control With Continuous-time Models and Applications to a Combustion- Engine Dynamometer". In: IFAC 12th Triennial World Congress, Sydney AUS, Proceedings. vol.10, pp.1-4.

Jones, W.M., Goetz, W.A., Canning, H., and de Voodg, A., 1988, "Closed Loop Fuel System and Low Emissions for a Natural Gas Engine". In: NGV-88 Conference, Sydney AUS. Proceedings. nº 38.

Kaiser, G., Zechnall, M., Plapp, G1-independence., 1988, "Closed Loop Control at Engine Management System Motronic". SAE paper 880135.

Lopes, J.A., 1996, "A Generalized Predictive Controller (GPC) Applied to the Air-Fuel Ratio Control Problem in Natural Gas Otto Cycle Engines, for Emission Reduction". (in Portuguese), MSc Thesis, São Paulo, 134 pp.

Moskwa, J.J., 1993, "Sliding Mode Control of Automotive Engines". Transactions of the ASME: Journal of Dynamic Systems, Measurement and Control, vol.115, pp.687-693.

Moscati, N.R., Trielli, M.A., Fleury, A.T., Nigro, F.E.B., and Lopes, J.A., 1996, "Modeling and Simulation of a Hydraulic Dynamometer". (in Portuguese) In: XI Congresso Brasileiro de Automática, São Paulo, SP. Anais. pp.1137-1142.

Moskwa, J.J., Hedrick, J.K., 1992, "Modeling and Validation of Automotive Engines for Control Algorithm Development". Transactions of the ASME: Journal of Dynamic Systems, Measurement and Control, vol.114, pp.278-285.

Noble, A.D., and Beaumont, A.J., 1991, "Control System for a Low Emissions Natural Gas Engine for Urban Vehicles". SAE paper 910255.

Powell, B.K., 1978, "A Simulation model of an Internal Combustion Engine-Dynamometer System". In: Summer Computer Simulation Conference, Newport Beach, CA. Proceedings. pp.464-473.

Schenk 1 "Hydraulic Dynamometer *DYNABAR*©, Publication" L3051/7 (in Portuguese), Schenck do Brasil, Ind. Com. Ltda.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2- 1997 - pp. 138-146

ISSN 0100-7386
Printed in Brazil

Stochastic Optimal Linear Parameter Estimation and Neural Nets Training in Systems Modeling

Atair Rios Neto
Universidade do Vale do Paraíba
Instituto de Pesquisa e Desenvolvimento
12245-720 São José dos Campos, SP Brasil

Abstract

Supervised training of feedforward neural networks for nonlinear mapping and dynamical systems modeling is addressed. Viewing neural nets training as a stochastic parameter estimation problem, results in Kalman filtering are adapted to develop training algorithms. Many levels of approximation are considered to develop a range of full non parallel to simplified parallel processing versions of algorithms, together with an adaptive approach intended to give to these algorithms the features of good numerical behavior and of distributing the extraction of learning information to all training data.
Keywords: *Neural Nets Supervised Training, Neural Nets and Systems Modeling, Stochastic Optimal Estimation Training Algorithms, Kalman Filtering Training Algorithms.*

Introduction

In recent years, the experience with optimal linear parameter estimation procedures has been explored to develop neural networks supervised training algorithms having the structure of recursive least squares (Chen and Billings, 1992) and of Kalman filtering (Singhal and Wu, 1989; Watanabe, Fukuda and Tzafestas, 1991; Scalero and Tepedelenlioglu, 1992; Chen and Ögmen, 1993; Chandran, 1994; Lange, 1995).

In this paper the author further explores the possibility given by Kalman filtering. Previously full non local processing (Rios Neto, 1994) and local parallel processing (Rios Neto, 1995) feedfoward neural nets training algorithms are presented together with the development of an adaptive procedure. Extending the stochastic optimal parameter estimation solution of the neural net supervised training problem one models the weight parameters as random walk stochastic processes. Noise dispersion adaptation (Rios Neto and Kuga, 1985) is then used as an automatic way of conditioning the covariance matrix of parameters estimation errors, thus avoiding loosing the capacity to extract information of new data as the processing goes on. The use of this adaptive procedure is thus intended to be effective along the processing in distributing to all data the extraction of learning information.

Fundamentals: Feedforward Neural Networks and Dynamic Systems Modelling

Among the types of feedforward artificial networks used for modelling and identification of systems(Chen and Billings,1992) the most basic and frequently used one is the Multilayer Perceptron made up of layers of basic artificial neurons connected forward, as illustrated in Fig.1, for the ith neuron of a kth hidden layer, with n_k neurons:

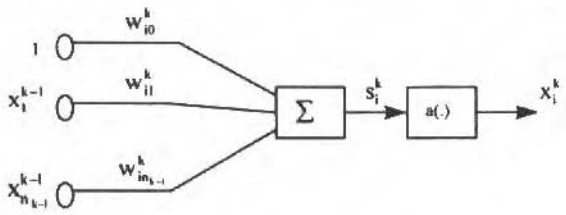


Fig. 1 Artificial Neuron

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

$$S_i^k = \sum_{j=1}^{n_{k-1}} w_{ij}^k x_j^{k-1} + w_{i0}^k \tag{1}$$

$$x_i^k = a(S_i^k),\ k = 1,2,\ldots,l-1 \tag{2}$$

with the activation function a(s) being typically taken as :

$$a(s) = 1/(1+\exp(-s)) \text{ or } a(s)=\tanh(s) \tag{3}$$

The inputs to the first hidden layer are $x_i^0 = x_i, i = 1,2,\ldots,n$, the network input vector. For the neurons of the output layer *(k=l)* it is sufficient to have and are frequently used zero threshold weights w_{i0}^l and identity activation functions:

$$\hat{y}_i \equiv x_i^l = \sum_{j=1}^{n_{l-1}} w_{ij}^l x_j^{l-1}, i = 1,2,\ldots,m \tag{4}$$

Feedforward artificial neural networks can be trained to uniformly and with the desired accuracy represent a nonlinear and continuous mapping (see, e.g., Zurada, 1992):

$$f \in C: x \in D \subset R^n \rightarrow y \in R^m \tag{5}$$

The theory already available (see, e.g., Hecht-Nielsen, 1990) guarantees that for the case of the Mutilayer Perceptron it is enough to have a neural network built with just one hidden layer , as illustrated in Fig. 2.

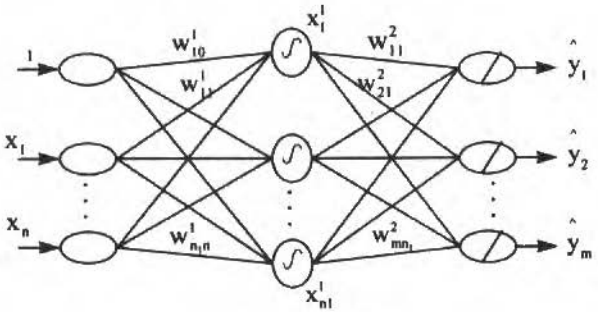


Fig. 2 One Hidden Layer Perceptron

The training of a feedforward network is usually done by supervised learning from mapping data sets:

$$\{(x(t), y(t)): y(t) = f(x(t)), t = 1,2,\ldots,L\} \tag{6}$$

adjusting (estimating) weight parameters to approximately fit the artificial neural net correspondent computational model to this data of input-output patterns.

The processing by the trained artificial neural net of the input data $x(t)$, to produce outputs $\hat{y}(t)$, can be viewed and treated as an parameterized mapping:

$$\hat{y}(t) = \hat{f}\left(x(t), w\right) \tag{7}$$

where w is the vector of weight parameters. In the case of the perceptron neural net with one hidden layer and hyperbolic tangent activation function (Fig.2), Eq. (7) is expressed as:

$$\hat{y}_i(t) = \sum_{j=1}^{n_l} w_{ij}^2 \left(tanh\left[\sum_{k=1}^{n} w_{jk}^1 x_k(t) + w_{j0}^1 \right] \right) \tag{8}$$

This capacity of feedforward artificial neural nets of representing nonlinear mappings can be used to approximately model dynamic systems of the type:

$$\dot{x} = f(x, u) \tag{9}$$

as long as $f(\cdot)$ is invariant in time (see, e.g., Chen and Billings, 1992). To do so , one has to implicitly assume that it is possible to consider a system as in Eq. (9) as if it was approximated by a discretized model like:

$$x(t+\Delta t) = \hat{f}\left(x(t), x(t-\Delta t), \ldots, x(t-n_x \Delta t); u(t), u(t-\Delta t), \ldots, u(t-n_u \Delta t), w\right) \tag{10}$$

which is of the type of Eq. (7) and where $n_x, n_u, \Delta t$ are to be adjusted together with the neural net arquitecture and size, depending upon the problem treated and desired accuracy. Notice that Δt can be treated as an extra component of the input to the neural net. What feedforward neural nets do in this case is to learn the mapping of Eq. (10).

At this point is opportune to remember the similar situation that occurs when numerical integrators are used and dynamic systems as in Eq. (9) are treated as discretized approximations , like in Eq. (10).

Supervised Training: Optimal Linear Estimation Procedure

A usual approach to solve the problem of supervised training of feedforward neural nets is to minimize, with respect to the vector of weights w, the functional:

$$J(w) = \frac{1}{2}\left[(w-\overline{w})^T \overline{P}^{-1} (w-\overline{w}) + \sum_{t=1}^{L} \left(y(t) - \hat{f}\left(x(t), w\right)\right)^T R^{-1}(t) \left(y(t) - \hat{f}\left(x(t), w\right)\right) \right] \tag{11}$$

given the input-output data $\{x(t), y(t): t=1,2,\ldots,L\}$, an a priori estimate $\overline{w}$, and the weight matrices $\overline{P}^{-1}, R^{-1}(t)$.

In the proposed solution (Rios Neto, 1994), a linear perturbation is adopted to approximate the functional of Eq. (11) in a typical ith iteration, imposing the condition that:

$$\alpha(i)\left[y(t) - \overline{y}(t,i)\right] \cong \hat{f}_w\left(x(t), \overline{w}(i)\right)\left[w(i) - \overline{w}(i)\right] \tag{12}$$

where, $i = 1,2,\ldots,I$; $\overline{w}(i)$ is the a priori estimate of w coming from the previous iteration, starting with $\overline{w}(1) = \overline{w}$; $\overline{y}(t,i) = \hat{f}\left(x(t), \overline{w}(i)\right)$; $\hat{f}_w\left(x(t), \overline{w}(i)\right)$ is the matrix of first partial derivatives with respect to w; and $0 < \alpha(i) \leq 1$ is a parameter to be adjusted in order to guarantee the hypothesis of linear perturbation. The resulting approximation of $J(w)$ in Eq. (11) is then:

$$J(w(i)) = \frac{1}{2}\left[(w(i)-\bar{w})^T \bar{P}^{-1}(w(i)-\bar{w}) + \sum_{t=1}^{L}(z(t,i)-H(t,i)w(i))^T R^{-1}(t)(z(t,i)-H(t,i)w(i))\right] \quad (13)$$

where the following compact notation was adopted:

$$z(t,i) \doteq \alpha(i)\left[y(t)-\bar{y}(t,i)\right] + \hat{f}_w(x(t),\bar{w}(i))\bar{w}(i) \quad (14)$$

$$H(t,i) \doteq \hat{f}_w(x(t),\bar{w}(i)) \quad (15)$$

The solution of minimizing the functional of Eq. (13) is formally equivalent (see, e.g., Jazwinski, 1970) to the following stochastic linear estimation problem:

$$\bar{w} = w(i) + \bar{e} \quad (16)$$

$$z(t,i) = H(t,i)w(i) + v(t) \quad (17)$$

$$E[\bar{e}] = 0,\ E[\bar{e}\bar{e}^T] = \bar{P} \quad (18)$$

$$E[v(t)] = 0,\ E[v(t)v^T(t)] = R(t) \quad (19)$$

with $\bar{e}$ and $v(t)$ not correlated and taken to have Gaussian distributions.

Proposed Off Line Solution

Following closely Rios Neto (1994), a Kalman filtering algorithm is proposed for an off line batch solution of problem of Eqs. (16) to (19):

$$\hat{w}(i) = \bar{w} + K(i)\left[z(i) - H(i)\bar{w}\right] \quad (20)$$

$$K(i) = \bar{P}H^T(i)\left[H(i)\bar{P}H^T(i) + R\right]^{-1} \quad (21)$$

$$\bar{w}(i+1) = \hat{w}(i),\ \alpha(i) \leftarrow \alpha(i+1) \quad (22)$$

where all the values of $t = 1,2,...,L$ were considered to define the extended vector $z(i)$, matrix $H(i)$ and the error vector v with covariance matrix R. The off line solution after iterations $i=1,2,...,I$ is given by:

$$\hat{w} = \hat{w}(I),\ P(I) = \left[I - K(I)H(I)\right]\bar{P} \quad (23)$$

If $\alpha(i)$ is sufficiently small to disregard high order terms in the linearization and enough redundancy exists in the training data ,then unless of bad numerical behavior of the Kalman filtering algorithm (Bierman, 1997) theory guarantees that $P(I)$ is an approximation for the estimation error covariance matrix, that is:

$$P(i) \cong E\left[(w-\hat{w})(w-\hat{w})^T\right]$$

For this off-line solution the following remarks apply:

i Since the natural situation is to have components of the error v uncorrelated , then R is diagonal and the recursive Kalman filtering algorithm can be used to process the vector $z(I)$ componentwize , avoiding the need of matrix inversion in Eq. (21) (Jazwinski, 1970);

ii If it happens that a new data set of pairs $\{x(t),y(t)\}$ is to be considered for network training, one only has to consider the recursive nature of Kalman filtering algorithm and take as new a priori information:

$$\overline{w} \leftarrow \hat{w} \qquad and \qquad \overline{P} \leftarrow P(I);$$

iii Though the backpropagation rule can be used to calculate the gradients in matrix $H(i)$ (Chandran, 1994), the algorithm presented does not attain paralell processing, and

iv Due to Kalman filtering typical behavior one should consider adopting some kind of factorization (Bierman, 1977) and or adaptive technique (Jazwinski, 1970; Rios Neto and Kuga, 1985) to avoid either numerical divergence or loosing the capacity of having learning distributed to all data.

Simplified Solutions: Paralell Processing

Algorithms with the structure of a Kalman filtering, which coincides with that of a recursive least squares, can be simplified to produce versions which preserve the local parallel processing capability of artificial neural networks (Chen and Billings, 1992). Exploring this possibility the author in a previous paper (Rios Neto, 1995) proposed approximated versions of the off line stochastic optimal parameter estimation algorithm (Eqs. (20) to (23)) and showed that even for the most simplified version of the stochastic optimal linear estimation, Kalman filtering algorithm leads to a local paralell processing algorithm still more general and sophisticated than the usual Backpropagation.

To better fulfill the purposes of this paper, in what follows one summarizes these results previously proposed by the author. Examining the equation bellow:

$$\alpha(i)\left[y(t)-\bar{y}(t,i)\right]=\hat{f}_w\left(x(t),\overline{w}(i)\right)\left[w(i)-\overline{w}(i)\right]+v(t) \qquad (24)$$

which is equivalent to Eq. (17), it can be seen that in the ith iteration the input-output data set can be locally processed to get an estimate of the vector of weights $w_{kl}(i)$ of the lth neuron in the kth layer if one considers that:

i For connection weight parameters $w_a(i)$ of neural net layers forward or after the one where processing is being done there are already available the estimated values $\hat{w}_a(i)$ and associated error $e_a(i)$ of known distribution;

ii For parameters $w_s(i)$, correspondent to weights of connections do the other neurons in the same layer, there is an a priori estimate $\overline{w}_s(i)$ with error $\overline{e}_s(i)$ of known distribution which can be taken as an approximation for w_s;

iii For the weight parameters $w_e(i)$ correspondent to connections to the neurons in earlier layers there is also an a priori estimate $\overline{w}_e(i)$ with error $\overline{e}_e(i)$ of known distribution which can be taken as an approximation for $w_e(i)$, and

With the previous assumptions, the problem of getting an estimate for the vector of weights $w_{kl}(i)$ of the lth neuron in the kth layer is reduced to the local estimation problem in the ith iteration and for $t=1,2,...,L$:

$$\overline{w}_{kl}(i) = w_{kl}(i) + \overline{e}_{kl} \tag{25}$$

$$\alpha(i)\left[y(t) - \bar{y}(t,i)\right] - \hat{f}_{w_a}\left(x(t), \overline{w}(i)\right)\left[\hat{w}_a(i) - \overline{w}_a(i)\right] = \hat{f}_{w_{kl}}\left(x(t), \overline{w}(i)\right)\left[w_{kl}(i) - \overline{w}_{kl}(i)\right] + \hat{f}_{w_a}\left(x(t), \overline{w}(i)\right)e_a(i) + \hat{f}_{w_s}\left(x(t), \overline{w}(i)\right)\overline{e}_s(i) + \hat{f}_{w_e}\left(x(t), \overline{w}(i)\right)\overline{e}_e(i) + v(t) \tag{26}$$

or, in a more compact notation, and for all the data corresponding to $t=1,2,\ldots,L$:

$$\overline{w}_{kl}(i) = w_{kl}(i) + \overline{e}_{kl} \tag{27}$$

$$\overline{z}_{kl}(i) = H_{kl}(i)w_{kl}(i) + \overline{v}_{kl}(i) \tag{28}$$

where $\overline{e}_{kl}$ is the correspondent partition of $\overline{e}$ in Eq. (16) and since in an ith iteration this problem can be locally and recursively solved with the Kalman filtering algorithm, starting with $\overline{P}$ diagonal, there results that the components of the errors $e_a(i), \overline{e}_s(i)$ *and* $\overline{e}_e(i)$ associated to parameters of different neurons are not correlated.

Further approximations can be done to produce simpler local parallel processing algorithms by:

iv Disregarding the off diagonal terms of the covariance matrix $R_{kl}(i)$ of the error $\overline{v}_{kl}(i)$ (Eq. (28)) allows to process $\overline{z}_{kl}(i)$ componentwize in Eq. (28), thus avoiding the need of matrix inversion; taking this approximation corresponds to consider $e_a(i), \overline{e}_s(i)$ *and* $\overline{e}_e(i)$ in Eq. (26) not correlated to each other and to *v(t), R(t)* diagonal, and to disregard off diagonal terms of the covariance matrices of:

$$\hat{f}_{w_a}\left(x(t), \overline{w}(i)\right)e_a(i), \hat{f}_{w_s}\left(x(t), \overline{w}(i)\right)\overline{e}_s(i), \hat{f}_{w_e}\left(x(t), \overline{w}(i)\right)\overline{e}_e(i);$$

v Disregarding the information on the level of accuracy in previous knowledge of $w_a(i), w_s(i)$ *and* $w_e(i)$, and taking these values as if they were:

$$w_a(i) = \overline{w}_a(i), w_s(i) = \overline{w}_s(i), w_e(i) = \overline{w}(i) \tag{29}$$

what implies a further simplified version of Eq. (26):

$$\alpha(i)\left[y(t) - \bar{y}(t,i)\right] = \hat{f}_{w_{kl}}\left(x(t), \overline{w}(i)\right)\left[w_{kl}(i) - \overline{w}_{kl}(i)\right] + v(t) \tag{30}$$

and which combined with Eq. (27) results in the simplified estimation problem for the data corresponding to $t=1,2,\ldots,L$

$$\overline{w}_{kl}(i) = w_{kl}(i) + \overline{e}_{kl} \tag{31}$$

$$z_{kl}(i) = H_{kl}(i)w_{kl}(i) + v \tag{32}$$

This last simplified version is still more sophisticated than the usual Backpropagation algorithm, and can be shown to be the result of application of Newton's method to the functional of Eq. (13) when the approximation of Eq. (29) is considered.

Adaptive Solution: Distributed Learning of Neural Weights

The problem with least squares type of estimators, and with Kalman filtering in particular , is that due to both algorithm bad numerical behavior and observation model errors, divergence usually occurs as many data sets are processed. This is due to the fact that the algorithm "learns too well the wrong information" (Jazwinski,1970) loosing capacity of keeping learning as new data are processed. What happens is an excessive and unrealistic decrease in the estimated dispersions of the errors in the calculated estimates. This corresponds to the situation of having the matrix of estimated covariances of the errors and the estimates with eigenvalues too close to zero.

To avoid this ill behavior and to try to keep a distributed and as much as possible uniform capacity of learning , it is common to use forgetting factor type techniques or more effective adaptive state estimation techniques like the one proposed by Jazwinski (1970) and modified by Rios Neto and Kuga (1985).

To apply an adaptive procedure based on a criterion of statistical consistency to balance a priori information priority with that of new learning information, the neurons connection weights parameters in the problem of neural net supervised training need to be modeled as random walk processes. Thus, in the ith iteration and for $t=1,2,\ldots,L-1$:

$$w(i,t+1)=w(i,t)+\eta(t) \tag{32}$$

$$E[\eta(t)]=0,\ E[\eta(t)\eta^T(\tau)]=Q(T)\delta_{t\tau} \tag{33}$$

where $\delta_{t\tau}$ is the Kronecker symbol and for the n_w weight parameters:

$$Q(t)=diag[q_j(t): j=1,2,\ldots,n_w] \tag{34}$$

With this modeling approximation for the neural weights, learning from the th input-output data pattern is transformed in the estimation problem:

$$w(i,t)=w(i,t)+\bar{e}(i,t) \tag{35}$$

$$z(t,i)=H(t,i)w(i,t)+v(t) \tag{36}$$

starting with $\bar{e}(i,1)=\bar{e},\ \bar{w}(i,1)=\bar{w}(i)$ and for $t=1,2,\ldots,L$..

To propagate estimates from t to t+1 Kalman filter predictor is used considering the dynamics of Eq. (32) :

$$\bar{w}(i,t+1)=\hat{w}(i,t) \tag{37}$$

$$P(i,t+1)=P(i,t)+Q(t) \tag{38}$$

where $P(i,t+1)=E[\bar{e}(i,t+1)\bar{e}^T(i,t+1)]$ *and* $P(i,t)$ is given by the filtering algorithm:

$$P(i,t)=\left[I-K(i,t)H(t,i)\right]\overline{P}(i,t) \tag{39}$$

where $\overline{P}(i,t)$ starts with $\overline{P}(i,1)=E\left[\overline{e}\overline{e}^T\right]=\overline{P}$.

The adaptation is done by adjusting the noise $\eta(t)$ dispersion such as to keep statistical consistency to attain distributed learning:

$$\beta E\left[v_j^2(t+1)\right]=H_j(t+1,i)\left[P(i,t)+Q(t)\right]H_j^T(t+1,i),\quad 1\le\beta\ge\beta_{lmx} \tag{40}$$

where $j=1,2,\ldots,m$ and β is to be adjusted close to 1 in order to have distributed learning.

This adaptive condition leads to the associated observation like condition , after some algebraic manipulations and adoption of a compact notation:

$$z^q(t+1,i,\beta)=H^q(t+1,i)q(t) \tag{41}$$

In order to use the same Kalman filtering algorithm the following associated estimation problem is considered:

$$0=q(t)+\overline{e}^q \tag{42}$$

$$z^q(t+1,i,\beta)=H^q(t+1,i)q(t)\ v^q(t+1) \tag{43}$$

$$E\left[\overline{e}^q\right]=0,\quad E\left[\overline{e}^q\overline{e}^{q^T}\right]=I_{n_w} \tag{44}$$

$$E\left[v^q(t+1)\right]=0,\quad E\left[v^q(t+1)v^q(t+1)\right]=R^q(t+1)=0 \tag{45}$$

which is a problem with exact observations that can be processed with Kalman filtering as long as one takes $Rq(t+1)$ in the limit as being zero (Freitas Pinto and Rios Neto, 1990). The solution gives a $\hat{q}(t)$ which is closest to zero in magnitude. Whenever a $\hat{q}_k(t)$ component is less than zero it is disregarded and taken to be zero , since the condition of positivity has to be observed.

Conclusions

Possibilities of results and past experience already existent in stochastic optimal linear parameter estimation were explored adapting Kalman filtering type of algorithms for feedforward neural networks supervised training . Full non parallel processing algorithms suitable for off line use as well as simplified parallel processing algorithms suitable for on line use which allows to stochastically treat the accuracy of training data were developed. Exploring past experience with state noise estimation in stochastic state observers, an automatic and adaptive approach was proposed which is expected to prevent these Kalman filtering based neural net weight estimators of loosing the capacity of distributing the extraction of information to all training data.

There is no reason for not expecting in neural nets training the same behavior stochastic optimal linear estimation algorithms have had in other applications of systems identification. The versions developed in this paper are all more sophisticated and realistic than the usual Backpropagation

algorithm. The price to be paid is more numerical complexity. This should be not a serious limitation for off line applications. For on line , real time applications , the computational resources already available for parallel processing may be enough to make competitive the use of the simplified versions , specially in mechanical systems where typical times of response are not so small.

References

Bierman,G.J., 1977, "Factorization Methods for Discrete Sequential Estimation", Academic Press, U.S.A.

Chandran, P.S., 1994, "Comments on Comparative Analysis of Backpropagation and the Extended Kalman Filter for Training Multilayer Perceptrons", IEEE Transactions on Pattern Analysis and Machine Inteligence, 16(8), pp.862-863.

Chen, G. and Ögmen, H., 1993, "Modified Extended Kalman Filtering for Supervised Learning", Int. J. Systems Sci.,24(6), pp. 1207-1214.

Chen, S. and Billings, S.A., 1992, "Neural Networks for Nonlinear Dynamic System Modelling and Identification", Int. J. Control, 56(2), pp. 319-346.

Freitas Pinto, R.L.U. and Rios Neto, A., 1990, "An Optimal Linear Estimation Approach to Solve Systems of Linear Algebraic Equations", J. Computational and Applied Mathematics, 33, pp. 261-268.

Iiguni,Y. and Sakai, H., 1992, "A Real -Time Learning Algorithm for a Multilayered Neural Network Based on the Extended Kalman Filter", IEEE Transactions on Signal Processing, 40(4), pp. 959-966.

Jazwinski, A.H., 1970, "Stochastic Processes and Filtering Theory", Academic Press, N.Y., U.S.A.

Lange, F., 1995, " Fast and Accurate Training of Multilayer Perceptrons Using an Extended Kalman Filter (EKFNet)", internal paper, DLR(German Institute Aerospace Research Establishment), Institute for Robotics and Systems Dynamics.

Rios Neto, A., 1994, "Stochastic Parameter Estimation Neural Nets Supervised Learning Approach", Proceedings. First Brazilian Congress in Neural Networks, Itajubá, Minas Gerais, Brasil, pp. 62-65.

Rios Neto, A.,1995, " Kalman Filtering Stochastic Optimal Estimation Algorithm and Usual Backpropagation in Neural Nets Training", Proceedings of Second Brazilian Congress in Neural Networks, Curitiba, Paraná, Brasil, pp. 139-144.

Rios Neto,A. and Kuga, H.K., 1985, "Kalman Filtering State Noise Adaptive Estimation", Proceedings of Second IASTED Int. Conference in Telecom and Control, TELECON'85, Rio de Janeiro, Brasil, pp. 210-213.

Scalero,R.S. and Tepedelenlioglu, N., 1992, "A Fast New Algorithm for Training Feedforward Neural Networks" , IEEE Transactions of Signal Processing, 40(1), pp. 202-210.

Singhal,S. and Wu, L., 1989, " Training Multilayer Perceptrons with the Extended Kalman Algorithm", In Advances in Neural Information Processing Systems, VI, Morgan Kaufman Pub. Inc., pp. 136-140.

Watanabe, K., Fukuda, T. and Tzafestas, S.G.,1991, "Learning Algorithms of Layered Neural Networks Via Extended Kalman Filters", Int. J. System Science, 22(4), pp. 753-768.

Zurada,J.M., 1992," Introduction to Artificial Neural System", West Pub. Co.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.1 - 1997 - pp. 147-153

ISSN 0100-7386
Printed in Brazil

Numerical Simulations of Interactions Among Aerodynamics, Structural Dynamics, and Control System

D. T. Mook
J. A. Luton
Virginia Polytechnic Institute and State University
Department of Engineering Science and Mechanics
24061-0219 Blacksburg, Virginia USA

Abstract

A numerical simulation of the interactions among the structure of an aircraft wing, the flow around it, and the devices that control the deflections of the ailerons is described. In the present simulation, the structure, flowing air, and controls are considered to be the elements of a single dynamic system. All of the governing equations are numerically integrated simultaneously and iteractively. The procedure is illustrated by an example of a very high-aspect-ratio, very flexible wing. Instead of a simple formula for the aerodynamic forces, there is s rather involved computer code. The input to this code, needed to impose the boundary conditions on the flowfield, is the velocity and position of all the points on the wing. As the airspeed increases and the angle of attack decreases in such a way that the lift force remains constant, the uncontrolled wing eventually begins to flutter. When the controls are turned on, the flutter can be suppressed up to approximately twice the critical airspeed.
***Keywords**: Flow Structure Interaction, Flight Control System.*

Introduction

During the last two decades or more, enormous effort has gone into the development of large computer codes that predict, rather accurately in a wide range of situations, unsteady hydro/aerodynamic forces on bodies executing general motions. In order to take advantage of the increased accuracy and generality that these codes offer to the development of control strategies, one must consider a rather different type of problem. In contrast with the familiar approach, in which the forces acting on a body are represented as somewhat simple functions (typically polynomials) of the state variables, the forces are obtained from a computer code; no direct formula exists. When the loads are required in the course of simulating the motion of the body, a computer code must be executed. Perhaps not surprisingly, this type of simulation promotes the concept of treating the complete dynamic system, but cannot make use of much of the vast amount of progress already made in the field of control of dynamic systems.

In the present paper, an example of this type of simulation is presented: the use of actively controlled motion of the ailerons of a very flexible, very high-aspect-ratio wing to suppress its flutter. The approach is to treat the flowing air, the elastic structure, and the devices moving the ailerons as elements of a single dynamic system. All the governing equations are solved interactively and simultaneously in the time domain. In the next section, the three elements are described; the manner in which all the equations are solved is presented after that, then some results are discussed, and finally some concluding remarks are made.

The Three Subsystems

The Structure

The long, slender wing is modeled as a prismatic, cantilever beam with constant properties. The governing equations written in dimensionless form are as follows:

$$\ddot{\phi} - D_1 \phi'' = q_1 Q_1 \tag{1}$$

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

$$\ddot{\upsilon} + D_b \upsilon^{iv} - D_b(2\upsilon'^2\upsilon'''' + 10\upsilon'\upsilon'''\upsilon'' + 3\upsilon''^3) = \frac{1}{2} q_b Q_b \upsilon'^2 + q_b Q_b - W \tag{2}$$

$$u = -\frac{1}{2}\int_0^y \upsilon'^2 dy \tag{3}$$

where overdots and primes represent derivatives with respect to time and position (y), respectively;

$$D_b = \frac{EI}{mL_c^2 V^2};\ D_t = \frac{GJ}{V^2 J_0};\ q_b = \frac{\rho L_c^2}{2m};\ q_t = \frac{\rho L_c}{2J_0};\ W = \frac{gL_c}{V^2};$$

EI is the flexural stiffness; GJ is the torsional stiffness; m is the mass per unit length of the wing; L_c is the characteristic length; V is the airspeed; J_0 is the mass polar moment of inertia per unit of length; ρ is the density of the air; w is the weight per unit length; ϕ is the angle of twist; υ is the deflection; and u is the axial displacement.

Here the nonlinear terms arise from expanding the curvative in the relationship between bending moment and deflection; there are other nonlinear terms that are formally the same order as these, but they are not included because the intent is to illustrate that for simulations of this type including nonlinear terms presents no difficulties.

In Equations (1) and (2) the aerodynamic twisting and bending moments per unit of length are represented by Q_t and Q_b, respectively. They are functions of $\upsilon, \dot{\upsilon}, \ddot{\upsilon}, \phi, \dot{\phi}$, and $\ddot{\phi}$ as well as the history of the motion. Moreover, they can only be obtained from a computer model of the flowfield. In the present model there is no structural coupling of twist and bending, but there is aerodynamic coupling. It is important to note that the acceleration appears on both sides of the equations of motion, explicitly on the left and implicitly on the right; there is no way to rearrange the equations so that the acceleration only appears on the left.

The boundary conditions on υ and ϕ have the following form:

$$\phi(0,t) = 0 \quad \phi'(L,t) = 0$$

$$\upsilon(0,t) = 0 \quad \upsilon''(L,t) = 0$$

$$\upsilon'(0,t) = 0 \quad \upsilon'''(L,t) = 0$$

To solve for the deflection of the wing, one must integrate Eqs. (1)-(3) while executing the computer code to determine Q_t and Q_b simultaneously.

The Flowing Air

The wings is modeled as a vortex sheet on the camber surface; the vortex sheet is represented by a grid of discrete vortex lines. Each elemental length of the grid is straight and has constant circulation, and hence, the velocity it induces is given by the Biot-Savart law. The circulations around the individual segments are determined by imposing the no-penetration condition at the centroid of each elemental area of the grid. At the wing tips and trailing edge, the difference between the pressures on the upper and lower surfaces must vanish; hence, vortex segments are shed from these edges into the flow, convect downstream at the local fluid-particle velocity while the circulation around any given segment remains constant, and form the wake. The flow at the surface of the wing and, therefore, the loads are affected by the vorticity in the wake. Because the vorticity in the wake now was shed earlier, the loads now are influenced by what happended earlier. Thus, the loads are history-dependent, and the historian is the wake.

The loads are obtained by integrating the pressure over the surface, and the pressure is obtained from Bernoulli's equation. For unsteady flows, Bernoulli's equation contains the derivative of the velocity potential. The potential depends on the velocity of the wing, among other things, and so its derivative is a function of the acceleration; hence, the acceleration appears on both sides of the equations of motion, explicitly on the left, implicitly on the right. The aerodynamic model is inherently nonlinear; moreover, it couples the flexural and torsional deflections through the loads.

The Control System

The motion of the wing is controlled by sending commands to a servomechanism that moves the aileron, which is located near the wing tip. In the present example, the control law that generates the commands has the following form:

$$\delta_c = G_1 \dot{\upsilon}(0.95L,t) + G_2 \dot{\phi}(0.95L,t) \tag{4}$$

where L is the span of the wing and G_1 and G_2 are constant gains. The two velocities can be obtained by integrating the signals from two accelerometers, one near the leading edge and the other near the trailing edge. The sum of the two signals is proportional to $\dot{\upsilon}$ and the difference is proportional to $\dot{\phi}$. The commanded flap deflection, δ_C, is fed to a severomechanism which moves the aileron and is governed by an equation of the following form:

$$\ddot{\delta} + 2\zeta\omega\dot{\zeta} + \omega^2\delta = \omega^2\delta_c \tag{5}$$

where δ is the actual deflection of the aileron and ζ and ω are constants. As in the case of the structural model, these equations do not have to be linear. The instantaneous hinge moment can be calculated as part of the simulation and could be included in the equation of motion.

Numerical Integration of the Governing Equations

The partial-differential equations governing the wing structure are integrated by first converting them to ordinary-differential equations: to this end we express the deflections as the following expansions:

$$\upsilon(y,t) = \sum_{n=1}^{M} B_m(t)\phi_m^F(y) \tag{6a}$$

$$\phi(y,t) = \sum_{n=1}^{N} C_n(t)\phi_m^T(y) \tag{6b}$$

where ϕ_m^F and ϕ_m^T are the linear free-vibration modes in flexure and torsion, respectively, for a cantilever beam. The time-dependent coefficients are the generalized coordinates of the motion. Expansions (6) are substituted into Eqs. (1) and (2) and Galerkin's procedure is followed to produce a set of second-order ordinary-differential equations. The integrals over the span of the wing of the products of the aerodynamic forces and modes produces the generalized forces. The second-order equations are written as a set of $2(M + N)$ first-order equations, which are integrated by Hamming's modified predictor-corrector method. This iterative algorithm was selected for two reasons: 1) the aerodynamic model functions best when the loads are evaluated at only integral time steps and 2) the accelerations appear on both sides of the equations of motion.

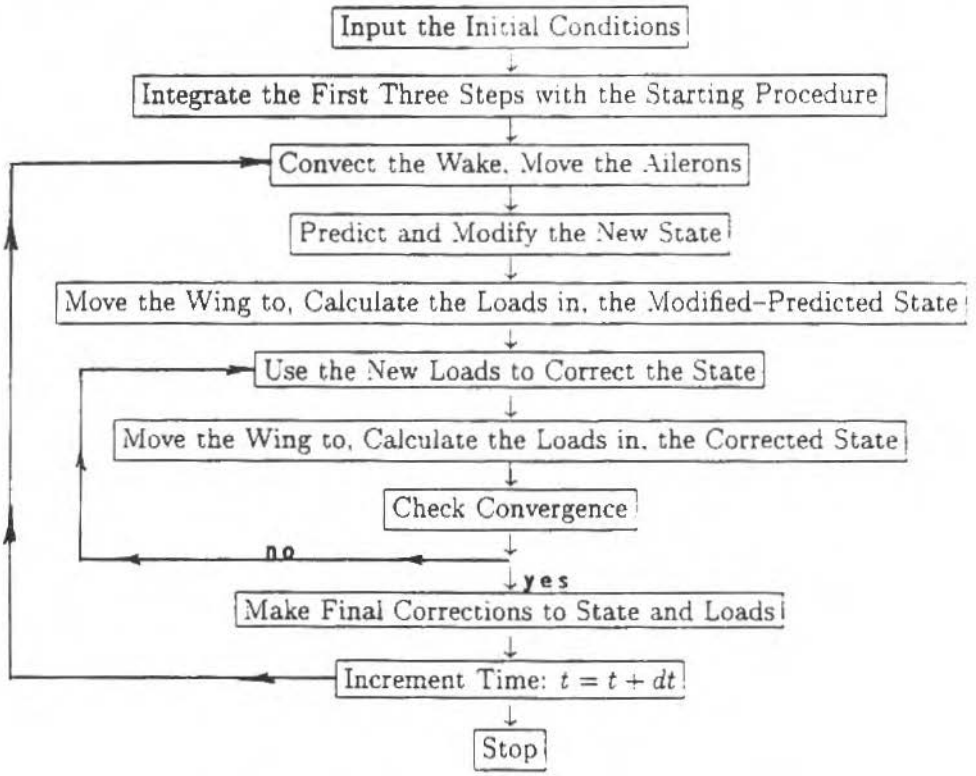


Fig. 1 Flow chart of the numerical scheme

The flow chart for the numerical scheme is shown in Fig. 1. The general algorithm is a four-step process. The solutions are required at the four previous time steps; hence, before the general algorithm is implemented, an Euler algorithm is used to generate the necessary previous solutions. Usually two, sometimes three, iterations are required to achieve convergence of the corrected state.

An Example

The following are the physical properties for the wing and air used in the present example:

Chord	1.0m
Aspect ratio	20.0
Elastic and inertial axis	0.305m from the leading edge
Mass per unit of span(m)	10Kg/m
Mass moment of inertial per unit span(J_0)	15.0Kg/m
Torsional stiffness(GJ)	2.493×10^5N - m^2

Bending stiffness(EI) $3.686 \times 10^5 N - m^2$

Density of the air in the freestream(ρ) $1.0572 Kg/m^3$

These values are representative of a high-altitude, long-endurance, unmanned aircraft.

The following are the natural frequencies (rad/sec) for the first few flexural (subscript b) and torsional modes:

$\omega_{b,1} = 675$ $\omega_{t,1} = 20.26$

$\omega_{b,2} = 42.30$ $\omega_{t,2} = 60.79$

$\omega_{b,3} = 118.45$ $\omega_{t,3} = 101.25$

$\omega_{b,4} = 383.75$

In the first simulations, the wing without control is run at 5 different speeds and five different angles of attack: 50 m/s, 5.9 deg; 75 m/s, 2.4 deg; 100 m/s, 1.2 deg; 125 m/s, 0.65 deg; 150 m/s, 0.35 deg. The lift is nearly the same for all cases; hence, the results correspond to the same aircraft (weight is constant) flying level at different speeds.

At each speed, the simulation is run until either a static steady-state emerges or flutter develops. The motion rapidly settles into a steady state at 50 m/s. At 75 m/s, the motion also settles into a steady state, but it takes at least four times longer than it took at 50 m/s; 75 m/s is slightly below the flutter speed. Although the model of the flow is inviscid, there is aerodynamic dynamic because of the phase between the loads and the motion. FFTs of the response reveal that the kinetic energy is concentrated at two frequencies: Those of the first flexural mode and the first torsional mode. These two frequencies change with speed; they are much closer at 75 m/s than they were at 50 m/s. At 100, 125, and 150 m/s the motion does not settle into a static steady state; instead its amplitude grows. FFTs of the motion reveal that now the kinetic energy is concentrated at a single frequency. At the onset of flutter this frequency is between the two frequencies that present in the decaying motion. As the airspeed increases and the flutter region is penetrated more deeply, the single frequency diminishes. Initially seven modes were included in the analysis, but numerical experiments revealed that only two are needed: the first flexural and torsional modes.

In Figure 2 the wing-tip deflection and rotation as well as the aileron deflection are plotted as functions of time with and without the control system being activated. The gains were obtained by trial and error. Clearly, an appropriately controlled aileron can suppress flutter, at least in the simulation. In this example, the airspeed is nearly twice the critical speed.

More details and references to other works can be found in Luton (1991).

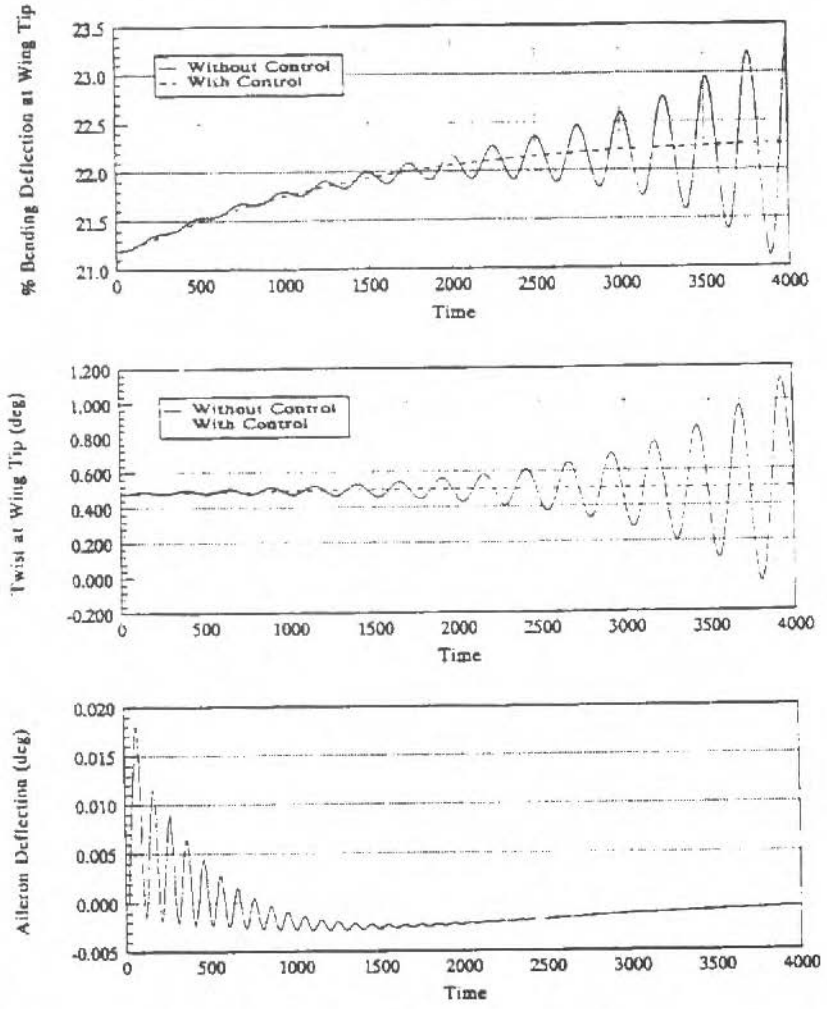


Fig. 2 The response of the wing with and without control. The velocity is 150 m/s and the gains are -1 and 1000.

Concluding Remarks

A new type of control problem has been illustrated in the present article. In this type of problem, the designer does not have a relatively simple formula to predict the loads, but must rely on a numerical model; every time the force is needed, a computer code must be executed. The numerical models offer greater accuracy and generality, but when they are used, much of the wealth of information developed in classical control theory does not apply. Hence, there is currently a need for some basic research in this area.

In the present simulations, interactions are virtually instantaneous. When a command to the controls is input, an unsteady response begins to develop. This response immediately affects the way the controls will perform, which in turn influences the developing response. In the present type of simulation, all of these interactions are modeled.

More development of computer models of flowfields will occur in the near future, and it is likely that more designs will be based on simulations like the one described here.

References

J.A. Luton, 1991, "Numerical Simulations of Subsonic Aeroelastic Behavior and Flutter Suppression by Active Control", Virginia Tech, Blacksburg, VA.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol XIX - No.2 - 1997 - pp. 154-163

ISSN 0100-7386
Printed in Brazil

Viscoelastic Neutralizers in Vibrations Abatement: A Nonlinear Optimization Approach

José João de Espíndola
Federal University of Santa Catarina
Mechanical Engineering Department
88040-900 Florianópolis, SC Brasil

C.A. Bavastri
National University of Comahue
Laboratory of Applied Mechanics
Neuquén, Neuquén Argentina

Abstratct

A general procedure for the optimization of the parameters of dynamic neutralizes is presented. It can be applied to the minimization of the vibration response and sound radiation of linear structures subjected to excitations in a specified frequency range. Modal theory and generalized equivalent quantity concept for the neutralizers, introduced by Espíndola and Silva (1992), are applied to a non-linear optimization scheme. The proposed procedure can be applied to relaxed and time invariant structures. It is not dependent on the structure complexity and the degree of discretization adopted. In such conditions, a significant reduction in computing work is achieved, if compared with the more traditional methods.
Keywords: *Vibration Isolation, Viscoelastic Neutralizers, Multidegree of Freedom Vibration.*

Introduction

Vibration neutralizers, often also called vibration absorbers, are mechanical devices to be attached to another mechanical system, or structure, called the primary system, with the purpose of reducing, or controlling vibrations and sound radiation.

Since neutralizers were first used to reduce rolling motions of ships (Den Hartog, 1956) many publications on the subject have steadily come to light, demonstrating their efficiency in mitigating vibrations and sound radiation in many structures and machines.

With modern technology of viscoelastic materials, which makes it possible to tailor a particular product to meet design specifications, vibration neutralizers are easy to make and apply to almost any complex structure.

In recent times, a great deal of effort has been done to extend and generalize the theory of vibration neutralizers, applied to more complex structures than the single degree of freedom undamped one, tackled by Den Hartog and Ormondroyd (1928).

Single degree of freedom neutralizers applied to particular positions of uniform beams, with particular boundary conditions, have been studied. Also mass distributed neutralizers have been analyzed. Simply supported uniform thin plates have also been considered as a primary system (Broch, 1946, Korenev and Reznikov, 1993).

In a recent work Espíndola and Silva (1992) derived a general theory for the optimum design of neutralizer systems, when applied to a most generic structure of any shape, any amount and distribution of damping. This approach has been applied to viscoelastic neutralizers of various types (Espíndola and Silva, 1992, Freitas and Espíndola, 1993).

The theory is based on the newly introduced concept of equivalent generalized quantities for the neutralizers. With this concept, is possible to write down the equations for the movement of the composite system (primary plus neutralizers) in terms of the generalized coordinates (degrees of freedom) previously chosen to describe the primary system alone, in spite of the fact the composite system has additional degrees of freedom.

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

This fact is crucial in the development of the theory. It permits a coordinate transformation using the modal matrix of the primary system, which is invariant during the optimization process.

In the modal space, it is possible to retain only few modal equations, encompassing the band of frequencies of interest. If coupling is not considered between these equations, then the neutralizer system can be designed to be optimum for a particular mode, in parallel with Den Hartog's simple optimization method. If a set of modal equation is retained, covering a particular frequency band, then a nonlinear optimization technique can be used to design the neutralizer system, to be optimum (in a certain sense) over that frequency band.

This paper reviews the concept of equivalent generalized quantities and how the composite system of equations can be written in terms of the generalized coordinates of the primary system only. It shows the coordinate transformation and describes the optimum design of neutralizers, in a frequency band, using a small set of modal equations. Numerical results are produced and discussed.

Nomenclature

C = ordinary viscous damping matrix of primary system

$\widetilde{c}$ = modified viscous damping matrix

$c_e(\Omega)$ = equivalent viscous damping constant of a neutralizer

c_j = j^{th} modal damping constant of primary system

$F(\Omega)$ = Fourier transform of f(t), or a transformed vector of applied loads

$\overline{G}\ (\theta,\Omega)$ = complex shear modulus of elastomer

$G\ (\theta,\Omega)$ = dynamic shear modulus of elastomer

i = $\sqrt{-1}$

K = ordinary stiffness matrix of the primary system

$\overline{K}\ (\theta,\Omega)$ = complex stiffness of a piece of rubberlike material

K_j = j^{th} modal stiffness of the primary system

L = constant dependent or the shape of the piece of rubberlike material (dimension L^{-1})

m = ordinary mass matrix of primary system

$\widetilde{m}$ = modified mass matrix

m_s = mass of a SDOF primary system

m_a = mass of a neutralizer

$M_a(\Omega)$ = dynamic mass at the root of a neutralizer

m_j = j^{th} modal mass of the primary system

$m_e(\Omega)$ = equivalent mass of a neutralizer

n = number of degrees of freedom of the primary system

$\hat{n}$ = number of modes in the frequency band of interest

p = number of neutralizers used for reducing vibrations and sound radiation

$p(t)$ = generalized principal coordinate of the primary system

$p(\Omega)$ = Fourier transform of $p(t)$

$q_j(t)$ = j^{th} physical generalized coordinate

$Q_j(\Omega)$ = Fourier transform of $q_j(t)$

$Q(\Omega)$ = Fourier transform of $q_j(t)$, or a vector of transformed displacement

$r(\Omega)$ = $G(\Omega) / G(\Omega_a)$

x = a vector of design variables

$Z_a(\Omega)$ = mechanical impedance at the root of a neutralizer

$a_{ks}(\Omega)$ = receptance function; response taken at q_k due to excitation associated with q_s

ε_a = Ω/Ω_a

$\eta(\theta,\Omega)$ = loss factor of elastomer

η_{aj} = loss factor of j_{th} neutralizer

θ = temperature

μ = m_a/m_s

$\mu_j = \dfrac{m_a \sum_{i=1}^{p} \phi_{k,j}^2}{m_j}$

Φ = matrix containing $\hat{n}$ eigenvectors of the primary system

Ω = angular frequency

Ω_a = anti-resonant frequency of a neutralizer

Ω_{aj} = anti-resonant frequency of j^{th} neutralizer

Ω_j = j^{th} natural frequency of primary system

$\{\}$ = column matrix, or vector

$[\,]$ = rectangular or square matrix

A^T = transpose of A

Equivalent Generalized Quantities for the Simple Neutralizer

The simple neutralizer has a single lump of mass connected to a based through a resilient device, assumed here of a viscoelastic nature (Fig. 1), with complex stiffness equal to (Espíndola, 1990):

$$\overline{K}(\theta,\Omega)= L\overline{G}(\theta,\Omega)= LG(\theta,\Omega)[1+i\eta(\theta,\Omega)] \tag{1}$$

For simplicity of notation, the letter θ, standing for temperature, will be omitted from now on.

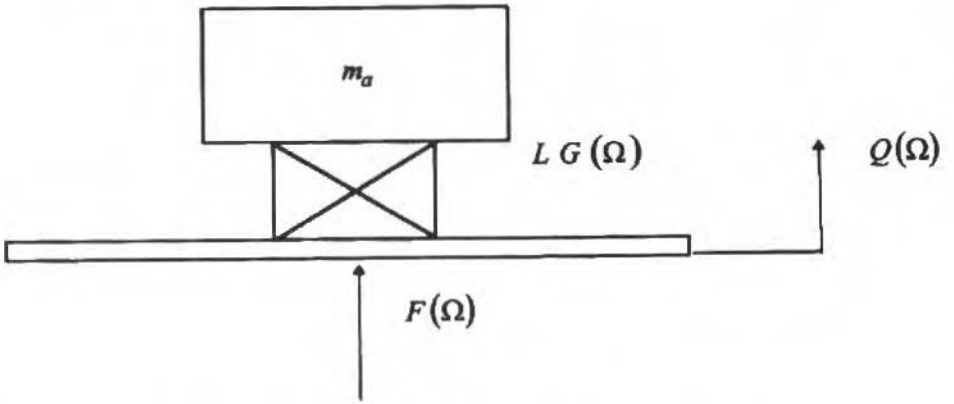


Fig. 1 A Simple Neutralizer Scheme

In the above figure $Q(\Omega)$ and $F(\Omega)$ stand for the Fourier transform of the massless base displacement $q(t)$ and the applied force $f(t)$, respectively.

It is a simple matter to verify that the impedance and the dynamic mass at the attachment (massless) plate are given by:

$$Z_a(\Omega)=\frac{-im_a\Omega L\overline{G}(\Omega)}{m_a\Omega^2 - L\overline{G}(\Omega)} \tag{2}$$

$$M_a(\Omega)=-m_a\frac{L\overline{G}(\Omega)}{m_a\Omega^2 - L\overline{G}(\Omega)} \tag{3}$$

The anti-resonant frequency of the simple neutralizer is defined as the one such that, in the absence of damping, makes the denominator of Eqs. (2) or (3) equal to zero:

$$\Omega_a^2=\frac{LG(\Omega_a)}{m_a} \tag{4}$$

(Note that, in absence of damping $\overline{G}(\Omega)=G(\Omega)$).

Since one can write $LG(\Omega) = LG(\Omega_a)r(\Omega)$, Eqs. (2) and (3) can be rewritten as:

$$Z_a(\Omega) = -im_a\Omega_a \frac{\varepsilon_a r(\Omega)\left[1+i\eta(\Omega)\right]}{\varepsilon_a^2 - r(\Omega)\left[1+i\eta(\Omega)\right]} \tag{5}$$

$$M_a(\Omega) = -m_a \frac{r(\Omega)\left[1+i\eta(\Omega)\right]}{\varepsilon_a^2 - r(\Omega)\left[1+i\eta(\Omega)\right]} \tag{6}$$

The equivalent generalized viscous damping is defined as the real part of the impedance (Eq. (5)) and for this simple neutralizer is:

$$c_e(\Omega) = m_a\Omega_a \frac{r(\Omega)\eta(\Omega)\varepsilon_a^3}{\left[\varepsilon_a^2 - r(\Omega)\right]^2 + \left[r(\Omega)\eta(\Omega)\right]^2} \tag{7}$$

In an analogous way, the equivalent generalized mass is the real part of Eq. (6):

$$m_e(\Omega) = -m_a \frac{r(\Omega)\left\{\varepsilon_a^2 - r(\Omega)\left[1+\eta^2(\Omega)\right]\right\}}{\left[\varepsilon_a^2 - r(\Omega)\right]^2 + \left[r(\Omega)\eta(\Omega)\right]^2} \tag{8}$$

Now, it is a simple thing to verify that both schemes shown in Fig. 2 are dynamically equivalent (Espíndola and Silva, 1992). The primary system "feels" the neutralizer as a mass $m_e(\Omega)$ attached to it along a generalized coordinate $q(t)$ and a viscous dashpot (even if the damping is solid) of constant $c_e(\Omega)$ linked to earth.

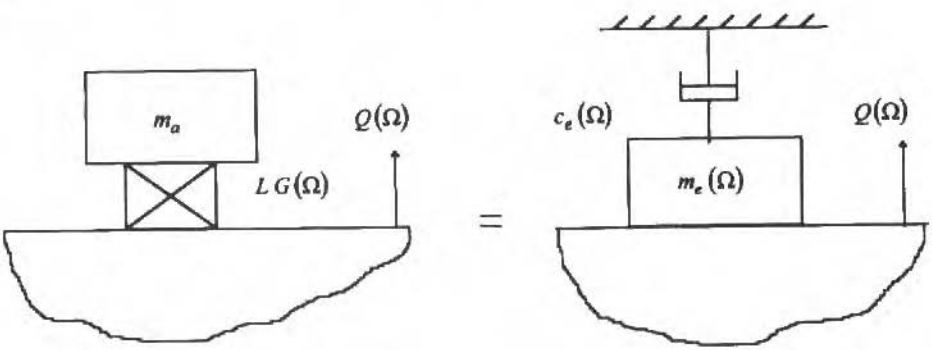


Fig. 2 Equivalent Systems

The dynamics of the resultant system (primary + neutralizer) can then be formulated in terms of the original physical generalized coordinates alone, although it has now added degrees of freedom. This is the main advantage of the concept of equivalent generalized quantities.

The above described generalized equivalent quantities correspond to a generic dynamic neutralizer. Particular cases, such as viscous or viscoelastic damping, can be analyzed by inserting the pertinent expression into the general equations.

If many such neutralizers are added, the equation of motion can be written as:

$$\left[-\Omega^2\widetilde{m}+i\Omega\widetilde{c}+K\right]Q(\Omega)=F(\Omega) \tag{9}$$

Note also that the effects to the added neutralizers are in modifying the mass and damping matrix. The vector of generalized coordinates of the primary system remain unchanged.

As a mean of illustration of the above ideas, assume that p simple neutralizers are attached along p physical generalized coordinates $q_{k1}, q_{k2}, \ldots, q_{kp}$. Their equivalent generalized masses and damping are $m_{e1}, m_{e2}, \ldots, m_{ep}$, and $C_{e1}, C_{e2}, \ldots, C_{ep}$.

The modified mass and damping matrices will be:

$$\widetilde{m}=m+\begin{bmatrix} 0 & & & & \\ & \ddots & & & \\ & & m_{e1} & & 0 \\ & & & \ddots & \\ & & 0 & & m_{ep} \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}_{n\,x\,n}$$

$$\widetilde{c}=c+\begin{bmatrix} 0 & & & & \\ & \ddots & & & \\ & & c_{e1} & & 0 \\ & & & \ddots & \\ & & 0 & & c_{ep} \\ & & & & & \ddots \\ & & & & & & 0 \end{bmatrix}_{n\,x\,n} \tag{10}$$

Now, in Eq. (9) assume the transformation:

$$Q(\Omega)=\Phi\,P(\Omega) \tag{11}$$

where Φ is the modal matrix of the primary system, obtained numerically or experimentally, and is of order $n\,x\,\hat{n}$, where n is its number of degrees of freedom and $\hat{n}$ is the number of eigenvectors actually computed, or measured. Normally $\hat{n}<<n$.

If Equation (11) is taken into Eq. (9), one get, assuming proportional damping in the primary system:

$$\left[-\Omega^2\left[diag(m_j)+m_A(\Omega)\right]+i\Omega\left[diag(c_j)+c_A(\Omega)\right]+diag(K_j)\right]P(\Omega)=N(\Omega) \tag{12}$$

where

$$N(\Omega)=\Phi^T F(\Omega), m_A(\Omega)=\Phi^T(\widetilde{m}-n)\Phi \text{ and } c_A(\Omega)=\Phi^T(\widetilde{c}-c)\Phi \tag{13}$$

Expression (12) represents a system of $\hat{n} << n$ equations and be solved directly with use of Eqs. (7) and (8). Returning back to (11), the solution in physical coordinates is accomplished.

From (12) and (13), it is easy to show that:

$$Q(\Omega) = \Phi D^{-1} \Phi^T F(\Omega) \tag{14}$$

where

$$D = D_0 - \Omega^2 m_A(\Omega) + i\Omega c_A(\Omega) \tag{15}$$

and

$$D_0 = diag\left(k_j - m_j \Omega^2 + i\Omega c_j\right) \tag{16}$$

From (14), the receptance matrix of the primary system, after the neutralizers have been attached, can be seen to be:

$$\alpha(\Omega) = \Phi D^{-1} \Phi^T \tag{17}$$

A particular member of this matrix is

$$\alpha_{ks}(\Omega) = \sum_{j=1}^{\bar{n}} \sum_{l=1}^{\bar{n}} C_{jl} \phi_{sl} \phi_{kj} \tag{18}$$

where C_{jl} are elements of $\left[\hat{D}\right]^{-1}$.

This can be compared with the receptances before the attachment of the neutralizers

$$\alpha_0(\Omega) = \Phi D_0^{-1} \Phi^T \tag{19}$$

and the pertinent response ratios computed:

$$R_{ks}(\Omega) = \frac{\alpha_{ks}(\Omega)}{\alpha_{0\,ks}(\Omega)} \tag{20}$$

The modulus of the response ratios can be taken as a measure of the efficiency of the neutralizers.

For systems with one degree of freedom, the recommended mass-ratio between neutralizer and primary structure, by Den Hartog (1956), is: $\mu = ma/m_s = 0.1$ to 0.25.

The use of the modal mass-ratio concept has been proposed by Espíndola and Silva (1992) for a system of multiple degree of freedom as:

$$\mu_j = \frac{m_a \sum_{i=1}^{p} \phi^2 \, k_i j}{m_j} \tag{21}$$

(Note that, if the eigenvectors are orthonormalized, $m_j = 1$)

Optimization for a Frequency Range

A technique for controlling the modes of interest, one at a time, is presented by Espíndola and Silva (1992). This assumes that matrices $m_A(\Omega)$ and $c_A(\Omega)$ are diagonal, which is not strictly true. This approach assumes μ_j given by Eq. (21) and uses the optimization technique suggested by Den Hartog (1956) and Snowdown (1959) for primary systems with one degree of freedom. For each mode, the corresponding equation in 12 is taken, neglecting any coupling.

This method gives pretty good results and, since $c_A(\Omega)$ is far from diagonal, the benefits of reducing response in one mode are generally spread to the neighboring ones.

In the optimization procedure of dynamic neutralizer presented here, neutralizer anti-resonant frequency and elastomer loss factor are considered design parameters while mass is fixed as expressed in Eq. (21). So, the design vector is:

$$x^T = (\Omega_{a1}, \Omega_{a2}, \ldots, \Omega_{ap}, C\eta_{a1}, C\eta_{a2}, \ldots, C\eta_{ap}) \tag{22}$$

the number of components being 2p.

The letter C, in front of the loss factors above, represents a fixed multiplicative factor, with the purpose of making $C\eta_{ai}$ of the same order of magnitude of the neutralizer's anti-resonant frequencies, numerically.

This is good for it speeds up convergence. A recommended value for C is given numerically by the average of the two limiting frequencies of the band of interest.

The initial guess for the neutralizer's mass is taken as follows: select a figure for μ_j within 0.10 to 0.25, say, and with expression (21) compute each neutralizer mass m_a and take the average in the band considered.

The objective function used in this work, in order to solve the optimization problem, is given by:

$$f(x) = \max_{\Omega 1 \le \Omega \le \Omega 2} \left| \alpha_{ks}(\Omega, x) \right| \tag{23}$$

where $\alpha_{ks}(\Omega x)$ is given in Eq. (18) for each design vector x, and $\Omega 1$, $\Omega 2$ are the lower and upper limits of the frequency band of interest.

The constraint functions could be defined from the following relation between the components of the design vector, x_1, and the range defined by x_i^L and x_i^U :

$$x_i^L \le x_i \le x_i^U , \qquad i = 1, 2p \tag{24}$$

After the optimization is completed, compute each neutralizer stiffness using the formula:

$$K_{ai} = m_a \Omega_{ai}^2 \tag{25}$$

where the m_a is the average neutralizer mass.

Now choose an elastomer (from a pertinent data sheet) whose maximum loss factor (at the environment temperature and frequency under consideration) is equal to η_{ai}. This maximum loss factor lies where the shear modulus increases approximately proportional to the frequency (see Fig. 3).

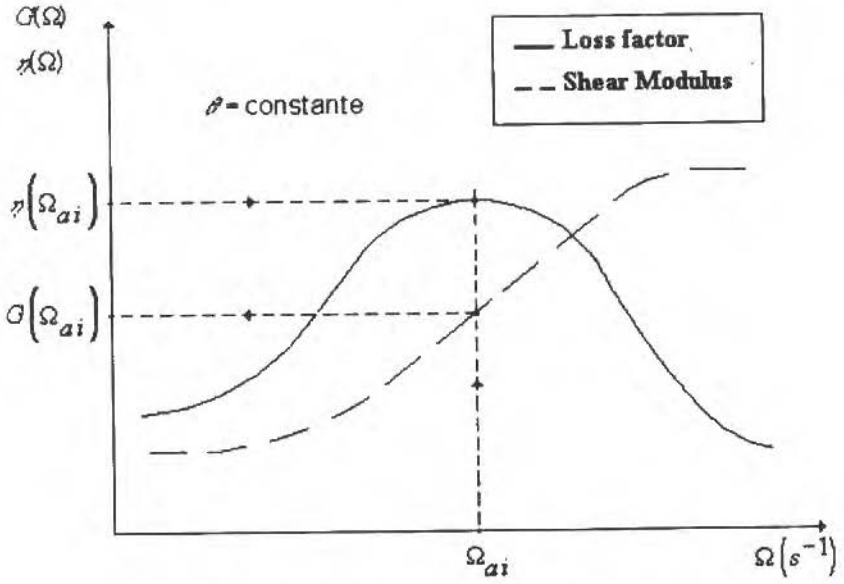


Fig. 3 Variation of η(Ω) and G(Ω) With Frequency

With η_{ai} (computed) and $G(\Omega_{ai})$ selected as in Fig. 3, one can compute the neutralizer's complex stiffness $\overline{K}(\theta,\Omega)$, provided L has been chosen properly (see formula 1). This ends the design process.

In practice, normally one material alone is selected, for simplicity and economy. So, in this case, an average loss factor and shear modulus is taken in the design.

Example

The above theory was applied to a freely supported steel rectangular plate.

The dimensions of the plate were 240 x 360 x 6 mm and it had a total mass of 4,0 Kg.

This plate has been divided in fifty-four elements and the modal parameter of the first eight modes were computed.

It was imagined that four neutralizers were fixed at the plate, one at each corner.

As explained previously, it was assumed that the resilient part of the neutralizers were made of a viscoelastic material, operating at the frequency transition zone, where the loss factor is large and the shear modulus is proportional to the frequency (see Fig. 3).

As an example of calculations, a band of frequencies from $\Omega 1 = 2875\ s^{-1}$ to $\Omega 2 = 4750\ s^{-1}$, comprising modes three and four, has been considered. An average mass for the neutralizers, computed as explained above, were 21.37 grams.

Table 1 shows the computed optimum values of anti-resonant frequencies for the neutralizers together with the corresponding loss factors.

Table 1 Optimum Values of Ω_{ai} and η_{ai} for Modes Three and Four

i	$\Omega_{ai}\ s^{-1}$	η_{ai}
1	3503	0.267
2	3413	0.304
3	4111	0.453
4	3403	0.356

In Fig. 4 a particular frequency response function is shown before and after the application of the neutralizers. These curves are produced to demonstrate the general effect of these simple devices.

Although the modes of concern were three and four, the coupling due to matrices $m_A(\Omega)$ and $c_A(\Omega)$ (see Eq. (12)) is beneficial to the neighboring modes.

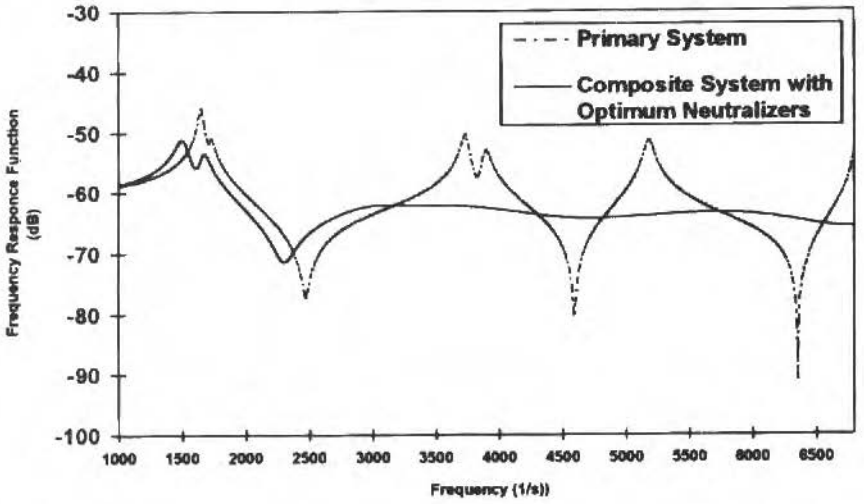


Fig. 4 A Particular FRF of the Primary Structure and the Effects of Neutralizers Upon it

One should note that the reductions in the frequency response levels are remarkable, with an addition of only 84,5 grams, i.e., of only two per cent of the primary plate mass.

Conclusions

The concept of equivalent generalized quantities can lead to a representation of the modal space of the composite system in terms of the modal parameters of the primary one. Retaining only a limited number of modal equations, an optimization scheme was devised, which leads to optimum parameter selection of a system of viscoelastic neutralizers.

This procedure is general, independent of the geometrical complexity of the linear primary structure, which is represented by its modal model. For best performance, the viscoelastic material should work in the frequency transition zone.

The performance of an optimum neutralizer system can be remarkable, at the expense of only a small increase in the overall weight, which makes this technique particularly valuable for light structures.

References

Alvarenga, E., 1995, "Optimization of Viscoelastic Pendulum Neutralizers", M.Sc. Thesis, UFSC, Brasil.

Boer, A., and Derksen, J., 1994, "Reduction of Noise Using Tuned Dampers", ISMA 19, Tools for Noise and Vibrations Analysis.

Broch, J.E., 1946, "A Note on the Damped Vibration Absorber", Journal Of Applied Mechanics, Trans. ASME, Vol. 68, pp. A284.

Den Hartog, J.P., 1956, "Mechanical Vibrations", 4th ed., McGraw-Hill, New York.

Espíndola, J.J., 1990, "Design and Performance of Viscoelastic Isolators and Suspensions for Aeronautic Engines", FBB Report.

Espíndola, J.J., and Silva, H.P., 1992, "Modal Reduction of Vibrations by Dynamic Neutralizers", International Modal Analysis Conference, San Diego, California.

Freitas, F.L., and Espíndola, J.J., 1993, "Noise and Vibration Reduction With Beam-Like Dynamic Neutralizers", 12th Brazilian Congress of Mechanical Engineering, Brasília, DF, pp. 1283-1286.

Jaquot, R.G., and Foster, J.E., 1977, "Optimal Cantilever Dynamic Vibration Absorber", Journal of Rng. For Industry, Trans. ASME, 138-141.

Jones, G.I.G., Nashif, A.D., and Stargardter, H., 1975, "Vibrating Beam Dampers for Reducing Vibrations in Gas Turbine Blades", Journal of Engineering for Power, 111-116.

Korenev, B.G., and Reznikov, L.M., 1993, "Dynamic Vibration Absorbers", John Wiley & Sons.

Nashif, A.D., and Jones, D.I.G., 1969, "A Resonant Beam Tuned Damping Device", Journal of Engineering for Power, 143-148.

Ordmondroyd, J., and Den Hartog, J.P., 1928, "The Theory of Dynamic Vibration Absorber", Journal of Applied Mechanics, Trans. ASME, Vol. 49-50, Part 1, pp. 1922.

Snowdon, J.C., 1966, "Vibration of Cantilever Beams to Which Dynamic Absorbers are Attached", Journal of the Acoustical Society of America, 1966, Vol. 39, N°. 5, Part 1, 878-881.

Snowdon, J.C., 1959, "Steady - State Behavior of the Dynamic Absorber", Journal of the Acoustical Society of America, Vol. 31, N°. 8, 1096-1103.

Snowdon, J.C., and Nobile, M.A., 1980, "Beamlike Dynamic Vibration Absorbers Acoustic", Vol. 44, 98-108.

Snowdon, J.C., Wolfe, A.A., and Kerlin, R.L., 1984, "The Crucciform Dynamic Vibration Absorber", Journal of Acoustical Society of America, 75(6), 1792-1799, June.

Snowdon, J.C., 1968, "Vibration and Shock in Damped Mechanical Systems", John Wiley & Sons, New York.

Young, D., 1952, "The Theory of Dynamic Vibration Absorbers for Beams", Proceedings, First U.S. National Congress of Applied Mechanics, ASME, New York, pp. 91-96.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 164-175

ISSN 0100-7386
Printed in Brazil

Malfunction Identification in Rotor Systems from Bearing Measurements Using Partial Models of the System

Nicolò Bachschmid
Riccardo Dellupi
Politecnico di Milano
Dipartimento di Meccanica
P.zza L. da Vinci, 32. 20133 Milano, Italy
bachschm@hp1 mecc.polimi it, dellupi@hp.1.mecc.polimi.it

Abstract

A method for the identification of different malfunctions which cause only or mainly 1x rev. vibration components is presented. The methodology is based on the model of the shaft alone, therefore avoiding the need of a linearized model of the oil film and of a reliable model of the casings and foundation, and uses the vibration readings in the bearings during coast-down transients. The results show that the method seems to be appropriate to distinguish between different causes such as concentrated unbalances, coupling misalignments and concentrated or distributed bows which could be produced by a partial rub (in a seal e.g.) or by a non uniform heating or cooling transient (in a generator or a steam turbine during load variations), and to determine the location along the rotor, the angular phase and the amount of unbalance or bow, in other words to identify the position and the severity of the malfunction.
Keywords: *Identification, Model Based Diagnosis, Rotordynamics.*

Introduction

Continuous monitoring of the shaft vibrations in correspondence of the bearings of main rotating machines, which generally are equipped with two proximity probes, and of the operating conditions of the plant, is standard practice in most power or petrochemical plants. A major change in the vibrational behaviour can be detected by comparing the actual data with previously recorded and stored data (when the rotor was free from any malfunction): these data should be related to the same operating conditions. Conventional European power plant turbogroups (with output powers in the range of 240 to 800 MW) are composed by several steam turbines whose shafts are rigidly coupled to each other and the last one of the set to a generator, and constitute therefore a 30 to 50 m long shaft, supported generally by 11 oil film bearings. Generally, at normal running speed, at least the same generator reactive and active output powers and the same temperatures (of steam, lubricating oil and cooling water e.g.) should indicate identical operating conditions. During a coast down transient, care should be taken to have the same initial conditions and the same stopping procedure (e.g. the vacuum breaking in the condenser should happen at the same time with respect to the starting point, and at the same turbine rotating speed). In these conditions, the vibrations of the shaft inside the bearings related to different periods of time, or to different run down transients, can be directly compared, avoiding the errors introduced by a certain thermal sensitivity of the complete system, rotor + oil film bearings + supporting structure, which leads to additional exciting causes (rotor thermal bow) and to a different frequency response (oil film coefficients depend on oil temperature and on alignment conditions).

The vibration vectors' changes in all bearings (which are the differences of the actual vibration vectors minus the original vibration vectors) in normal operating conditions and at the different rotating speeds during the coast down can then be analyzed in order to assess a "symptom matrix" in which the characteristics of the vibrations are pointed out (such as frequency content, amount of different components and corresponding location of bearing, and so on). This symptom matrix can be compared with a "cause matrix", and a first guess of type of malfunction and its rough location (in which shaft e.g.) can be determined. This type of approach or a decision tree approach have been applied in some advanced monitoring and diagnostic systems (see e.g. Kanki et al. (1993) and Tanaka (1993)).

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

The nest step in diagnostics is a model based diagnostic procedure which has the advantage of identifying not only the type of malfunction but also its amount or severity and its location along the rotor system. This step can be called malfunction identification. The model based diagnostic procedure is based essentially on the comparison of the actual vibrational behaviour (obtained as difference as previously described) with the calculated behaviour, which is generally obtained through a model of the complete system (rotor + bearings + casing or pedestals + foundation + a suitable model of the considered malfunction). In order to obtain acceptable results the model has to be sufficiently reliable. And to allow to operate in the frequency domain and to have the system represented by its elastodynamic matrix which has to be inverted, the model must be linear: therefore linearized oil film stiffness and damping coefficients, as usual in rotordynamic calculations, have to be used.

In the following paragraphs first the obtainable results with the complete linear model of the system are discussed, and then the possible errors in the identification procedure due to the linear model of the oil film bearings and to a poor model of the supporting structure are emphasized. In the last section then a modification of the identification procedure is presented which avoids the linearization of the oil film forces and works without any model of the supporting structure, using the "partial" model of the shaft alone.

Applications of model based malfunction identification

This method, which uses the linear model of the complete system, furnishes excellent results when applied to simulated data, as shown e.g. in Bachschmid et al. (1995a), but if experimental data from real machines in normal operating conditions are used, then acceptable results are obtained only if the data are somehow filtered (if e.g. some of the available measured bearing vibrations are disregarded and/or some speed ranges are disregarded) as reported in Bachschmid et al. (1996). A good agreement between experimental and calculated results is found only in a smaller speed range and generally by tuning previously the model. Many different reasons may be responsible for this situation:

- The run down transients which are compared might be carried out in slightly different thermal situations in which the rotor exhibits different thermal bows. This effect could not be taken into account in the mentioned papers since the records of the values of the parameters which characterize the thermal situation of a rotor, just before and during the run down transient, were not available;
- The foundation of the machine, which can strongly influence the dynamical behaviour of the rotor, is poorly represented by one d.o.f mass damper and spring systems (as usual considered in many rotordynamic calculation programs). Unfortunately, in the above mentioned case studies, a more reliable model of the supporting structure was not available, and
- The oil film may be strongly non-linear, and is poorly represented by the linearized model.

In order to show which are the non linear effects in the oil film bearing which may affect the identification procedure based on the linearized model, in the following paragraph some calculated results with the non linear model are presented.

An in order to remind how important the dynamic effects of the foundations are, especially in the case of flexible foundations, which sometimes are used in place of the more massive concrete foundations, some calculated results related to a four oil film bearing test rig on a flexible supported foundation are shown in the next paragraph.

The method which is presented in the last section of this paper has the goal to overcome the last two problems: the model of the foundation is not anymore needed and the linearized model of the oil film is substituted by the non linear oil film forces. The effectiveness of the methodology will be tested with numerical results obtained on the above said test rig model shown in Fig. 1.

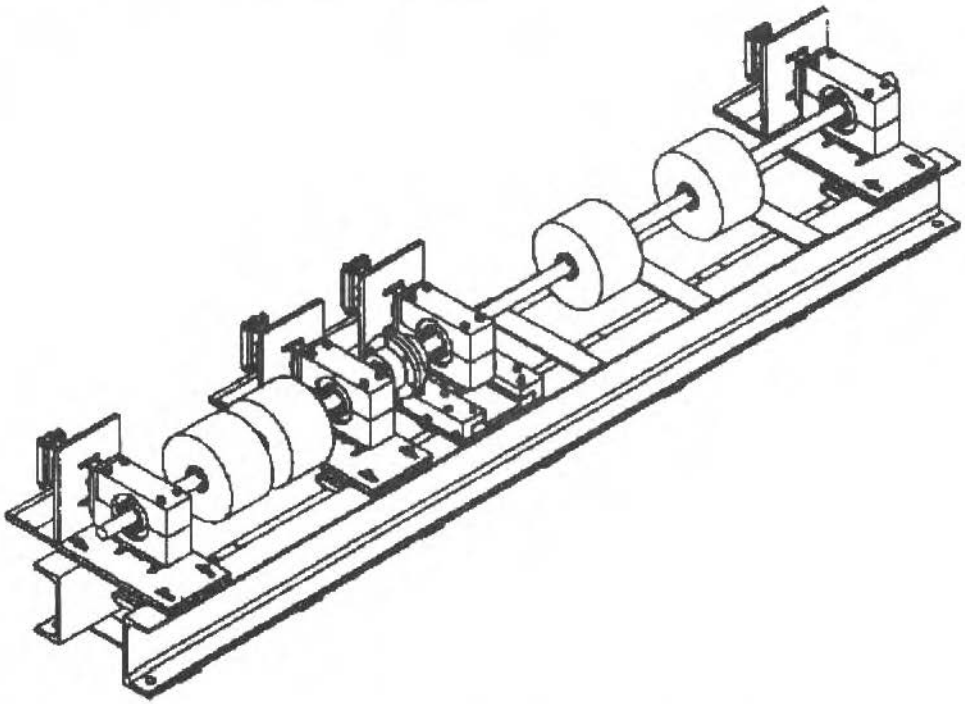


Fig. 1 The Test Rig Composed by two Rigidly Coupled Shafts Supported by 4 Oil Film Bearings Equipped With Force Transducers Mounted on a Flexible Supported Foundation Frame

Some causes produce only major changes in 1x rev. components: these are mainly a concentrated unbalance (caused e.g. by a blade loss) and a permanent or transient thermal bow. The bow could be due to an angular coupling misalignment when it is located close to the coupling in between two bearings, or to a rub in a seal (which causes a local unsymmetric and therefore the bow) when it is located close to a seal, or could be distributed, due to a non-symmetric heating or cooling of the central body of the rotor (which in a generator can be caused by a wire short circuit or a cooling duct obstruction, and in a turbine by a non uniform steam distribution). Other causes produce generally also some changes in other components: a crack produces changes in 2x and 3x components, a bearing misalignment produces also changes in the mean static component, a loosening part produces higher harmonics and so on.

It is difficult to separate or distinguish between the different causes which produce only 1x components. This is also shown in Bachschmid et al. (1996). But in order to test the effectiveness of the proposed method, only 1x causes are considered.

Non-linear Oil Film Effects

Some non-linear effects in steady state operating conditions of oil film bearings are shown by means of numerical results obtained by time step integration of the equation of motion of the finite element model of a fairly simple rotor (show in Fig. 2) supported by two 2-lobe "lemon" shape oil film bearings. The weight and the lengh of the shaft and the type and dimensions of the bearings are typical for a H. P. steam turbine. In each time step the oil film forces are calculated by integrating the Reynolds equations.

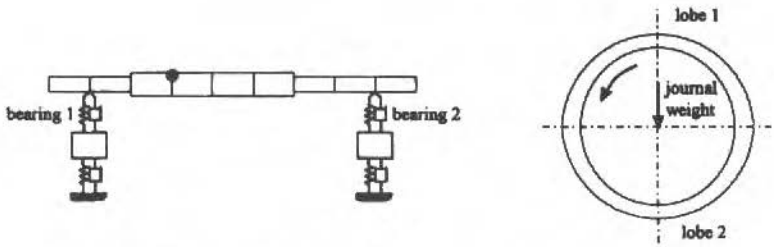


Fig. 2 The HP Steam Turbine Rotor FEM Model and Schematic of the "Lemon" Bearings

Two different unbalances (of 0.1 kg and 0.5 kg) have been applied to a node of the rotor and, after some 20 to 50 complete revolutions of the shaft, the steady state situations shown in following figures have been reached. In Fig. 3a and 4a the orbits of the journal centres at 2250 rpm in bearing 1 and bearing 2 due to the small unbalance, and Fig. 3b and 4b the corresponding orbits due to the high unbalance, inside the loci of the different positions allowed by the bearing clearance, are represented. The running speeds is close to the second critical speed of the shaft, therefore the high unbalance orbit in bearing 1 is so large.

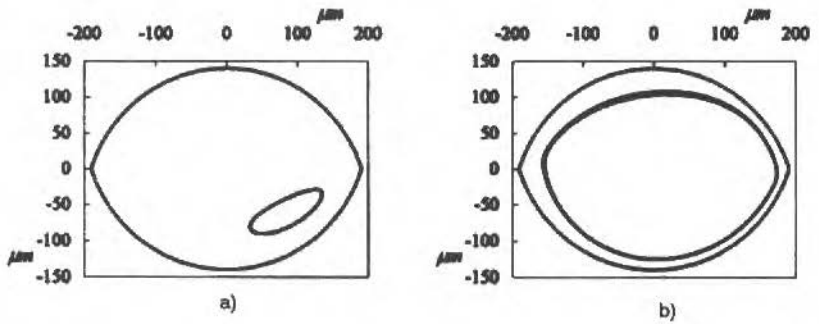


Fig. 3 Orbits of Shaft Journal in Bearing 1 at 2250 rpm: a) due to unbalance of 0.1 kg, b) due to unbalance of 0.5 kg

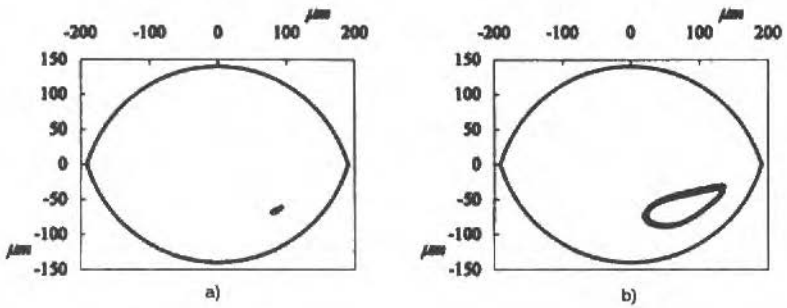


Fig. 4 Orbits of shaft Journal in Bearing 2 at 2250 rpm: a) due to unbalance of 0.1 kg, b) due to unbalance of 0.5 kg

With a linear behaviour, we would find the high unbalance orbit with the same shape of the small unbalance orbit, but enlarged 5 times. It is evident from the figures that the behaviour is different, the size and the shape of the orbits are quite different, and in the Fourier analysis of the orbits we would

find differences in amplitude of the 1x revolution component, and also higher harmonics. Similar results are found for bearing 2 in which the journal vibration amplitude is much smaller, and in which therefore also the non-linear effect should be negligible. In the same bearing 2 the orbits at a different rotating speed (2000 rpm) are represented in Fig. 5a (for the small unbalance) and in Fig. 5b (for the large unbalance), in the different scales: the shape of the orbits, and consequently the spectral components, are slightly different in the two situations as a result of the non-linear behaviour.

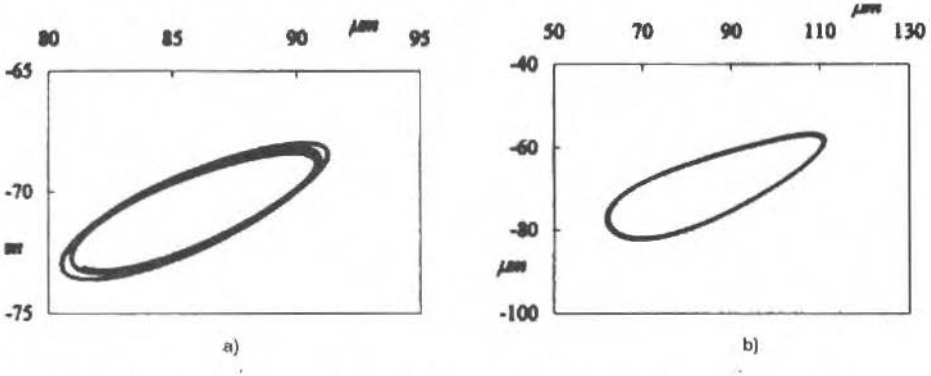


Fig. 5 Orbits of the Shaft Journal in Bearing 2 at 2000 rpm: a) due to unbalance of 0.1 kg, b) due to unbalance of 0.5 kg

If we would use these orbits as measured quantities for identifying the exciting forces on the rotor with a linearized model, we would have considerable errors in determining the type of exciting force (since also 2x rev. components are presented), its location and its amount, because the linear model is not able to reproduce adequately the real non-linear behaviour.

A more complete analysis of the non linear effects of oil film bearings in steady state operating conditions is reported in Bachschmid and Dellupi (1997).

Foundation Effects

The supporting structure effects on the dynamical behaviour of the rotor are well known, and many papers have been published on this topic, proposing also different methods for calculating the behaviour of the rotor coupled to the supporting structure by means of the oil film bearings. Since the models of the supporting structures are often rather complicated and unreliable in the results, many efforts have been made for developing model updating techniques and model identification techniques (see e.g. recently Friswell et al. (1996) and Vania (1996)) by means of some experimental results, in order to obtain more reliable models which are sometimes so important for calculating the true behaviour of the rotor on its supporting structure.

The effects of the supporting structure dynamical behaviour on the rotor consist in shifting of some critical speeds, sometimes in nearly suppressing a critical speed, in generation the so - called "foundation critical speeds" and, in general, in significantly changing the amplitude and phase of the frequency response curves. These effects will be shown by means of numerical results obtained with the model of the reduced scale test rig (represented in Fig. 1).

In Figure 6a the unbalance response curves in vertical direction of the node where the unbalance was applied is represented for the case of flexible foundation (continuous line) compared with the case of rigid foundation (dashed line) and in Fig. 6b the same curves in horizontal direction are represented. The total mass of the foundation is 136 kg and the rotating mass is 40 kg. The ratio of the rotating mass may be relevant in determining the influence of the foundation on the dynamical behaviour of the rotor: the test rig might be representative only for turbogroups on light foundations. More massive foundations would have with the same excitation smaller vibration amplitudes.

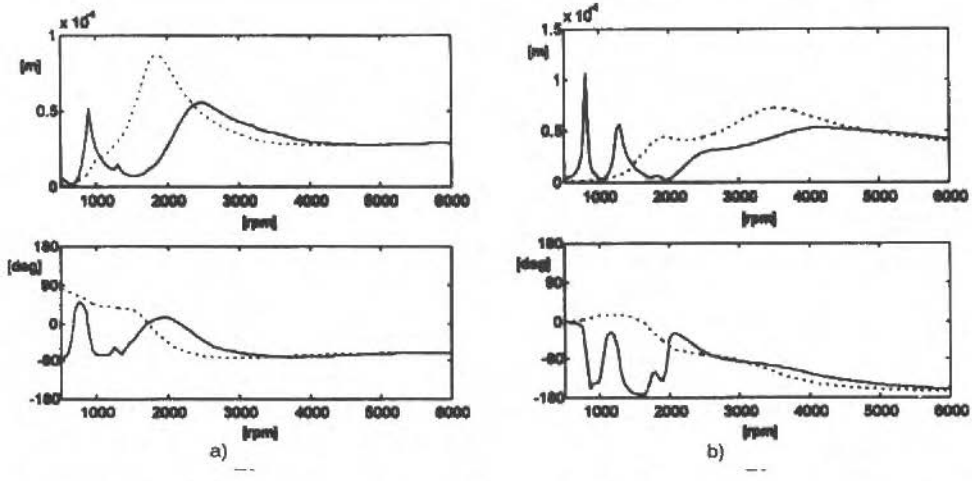


Fig. 6 Unbalance (0.001 kg at Test Rig's 1st disk) Response Curves at 1st Disk of the Test Rig Rotor on Rigid Foundation (Dashed Line) and Flexible Foundation (Continuous Line), in Amplitude and Phase: a) in Horizontal Direction, b) in Vertical Direction

The above diagrams show clearly that the dynamical behaviour of the supporting structure has a great importance in determining the response of the rotor, at least in the examined case. This emphasizes the need of a reliable model of the supporting structure, if the complete model of the system is used for the force identification procedure, or stresses the advantage of a method which avoids completely the use of a model of the supporting structure. This last condition is fulfilled if the proposed methodology is used: in this case however also the absolute vibrations of the bearing housings have to be measured, in addition to the relative displacements of the shaft inside the bearings which are measured by the proximity probes.

Description of the Identification Method

With two proximity probes in each bearing the orbits of the shaft inside the bearings can be determined. The geometry of the bearings and its clearances are known. By solving the Reynolds equations in different points of the orbit, the horizontal and vertical oil film forces can be calculated in each point. Generally monitoring systems acquire 8 or 16 points of each orbit during one revolution. These forces are, if steady state motion has been reached, 1x revolution periodic and can be analyzed by Fourier series, obtaining the mean static force component and several harmonic components. Using a harmonic balance concept (which is possible because of the rotor model linearity) the first harmonic component (1xrev. component) in the bearings should balance the 1xrev. exciting forces and the 1xrev. inertia forces acting along the shaft due to the vibrations. Therefore we can consider the unconstrained free-free rotor, as shown in Fig. 7, on which in correspondence of the bearings the known 1xrev. oil film force components (indicates by F_e) are applied and, in an unknown position, the unknown exciting force (indicated by F_i) is applied:

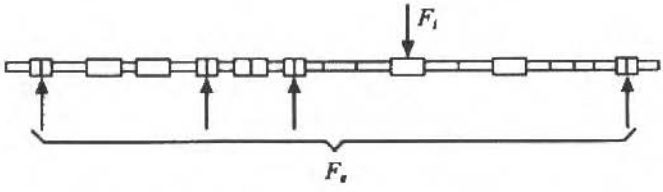


Fig. 7 Schematic Diagram of "Forces" Method

If the absolute vibrations of the shaft in correspondence of the bearings are known, a standard force identification procedure based on least square fitting of the vibrational data (as proposed by Bachschmid et al. (1996)) can be used for identifying the exciting forces moduli and phases and their position along the shaft. A similar approach, but limited to a rigid supporting structure and with a deterministic method instead of a least square approach is proposed by Krodkiewski et al. (1993).

The absolute vibrations of the shaft can be obtained by adding to the relative vibrations of the shaft inside the bearings (which are measured by two proximity probes) the absolute vibrations of the bearing, which can easily be measured by a couple of accelerometers fixed to each bearing housing.

Many different exciting causes can be represented by a system of equivalent forces or couples: besides the unbalance, also the concentrated or distributed bow can be represented by two equal and opposite couples which are applied to the extremity nodes of the bowed shaft portion, which can be one element of the f.e. model of the shaft in case of coupling misalignment (as proposed by Bachschmid et al. (1995c)) or several elements of the central body of the rotor in case of a thermal bow (as shown by Bachschmid and Del Vescovo (1990)). Also a transverse crack can be modeled by means of an equivalent system of forces and couples: with this model e.g. the location and the depth of a crack have been identified by applying the least square fitting procedure to 1xrev. and 2xrev. deflection values measured in 4 different sections along the rotor during slow rotation on a lathe (as reported by Lapini et al. (1993)).

The main differences with respect to the procedure which uses a complete linear model of the system (rotor + oil film + supporting structure) is that the forces are applied to an unconstrained free-free rotor, which exhibits high vibration amplitudes (also at bearing locations) if the external force balance of exciting force and bearing forces is not exactly fulfilled. The procedure might therefore be more sensitive to force identification errors than the least square fitting procedure applied to a constrained rotor, and this might result in both advantages (higher precision) and disadvantages (possible instability); but these effects have not been recognized in the numerical simulations.

Mathematical Formulation of the Method

In each time instant the relative position of the shaft journal with respect to the bearing is measured and the corresponding velocity can be calculated in order to evaluate the oil film forces F_e with a suitable model of the oil film. These forces are periodic (in steady-state conditions), and can be represented by a Fourier series:

$$Fe = F_{e0} + F_{e1}e^{i\Omega t} + \ldots + F_{en}e^{in\Omega t} \tag{1}$$

where F_{e0} is the mean static component and F_{e1} the 1st harmonic component. All forces are split in horizontal and vertical complex components.

Also the vibrations are periodic: the first harmonic component of the absolute vibrations of the shaft in correspondence of the bearings X_e is given by:

$$X_e = X_{er} + X_{eb} \tag{2}$$

where X_{er} is the relative vibration of the shaft journals (with respect to the bearing), measured by the proximity probes, and X_{eb} is the absolute vibration of the bearing housing, measured by the accelerometers. These vibrations are split in horizontal and vertical complex components.

In the frequency domain referring to the 1st harmonic components following force balance equations holds:

$$(-\Omega^2 M + i\Omega R + K)X = F_e + F_i \tag{3}$$

where M, R and K are the mass internal damping and stiffness matrices respectively of the shaft alone, X is the vector of the displacements of all nodes (four complex components in each node), and F_i are the unknown exciting forces.

$$F_i = \Omega^2 U + B \tag{4}$$

U represents the unbalance vector (in case of concentrated unbalance only two elements are non zero) and B the bow vector (composed by two moments in each one of the extremity nodes of the interested shaft portion).

The dynamic rotor matrix can be inverted:

$$a = (-\Omega^2 M + i\Omega R + K)^{-1} \tag{5}$$

and the a matrix can be partitioned separating the "internal" displacements X_i of the nodes of the rotor from the "external" displacement of the shaft journals X_e:

$$a_{ii}F_i + a_{ie}F_e = X_i \tag{6}$$

$$a_{ei}F_i + a_{ee}F_e = X_e$$

The second set of equations of (6) can be used for determining the unknown F_i vector. The elements of F_i are generally less than twice the number of the bearings multiplied by the number of different rotating speeds so that a least square approach can be used for determining F_i:

$$((X_e - a_{ee}F_e) - a_{ei}F_i)' ((X_e - a_{ee}F_e) - a_{ei}F_i) = min \tag{7}$$

The sum is extended to all the considered different rotating speeds.

A residual may be defined by the ratio of left hand side of Eq. (7), divided by the sum of squared values of the known term $(X_e - a_{ee}F_e)$. The residual is calculated for each possible position(s) or node(s) of application of the exciting cause, calculating subsequently the effect of different causes (U or B). The lower the residual is, the higher is the accordance of measured vibrations with the calculated ones. The lowest obtained minimum identifies the exciting cause. The most probable location of the exciting cause is there where the residual reaches its minimum: the corresponding value of F_i furnishes its modulus and phase.

Validation of the Method by Means of Numerical Results

The described identification method was applied to test rig rotor model, simulating numerically its behaviour. One unbalance (or 2 different unbalances) or one bow have been applied to the rotor, and the corresponding frequency response curves ate the 4 bearings were obtained. The relative displacements and velocities of the shaft journal with respect to the bearing allow to calculate the bearing forces, which were then used together with the absolute vibrations of the shaft in the identification procedure. Always the same 12 equally spaced running speeds in the range of 500 - 6000 rpm have been considered.

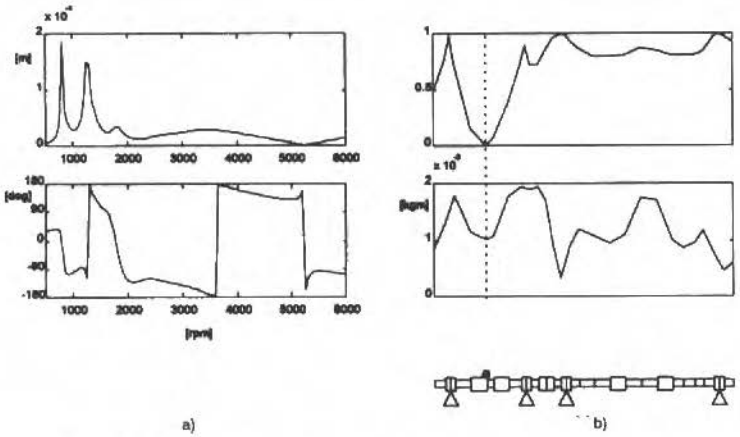


Fig. 8 Unbalance (0.001 kg) on Test Rig's 1st Disk: a) Frequency Response at Bearing 1 (Journal's Vertical Displacement); b) Residual and Identified Unbalance Amount

Figure 8a shows the calculated frequency response curve of bearing 1 due to an unbalanced 1st disk, and in Figs. 8b the value of the residual versus the portion along the shaft, and the corresponding identified unbalance amount and phase are represented. The position, amount and phase are identified with excellent accuracy.

In Figure 9a the calculated frequency response curves in the same bearing due to 2 different unbalances placed on disks 1 and 3 (on the 2 different shafts), and in Fig. 9b the residual showing 2 different minima (the lower near the first unbalance location) are represented. In Fig. 9c the residuals corresponding to the second unbalance in its different positions along the shaft, maintaining for the first unbalance the position of the previous absolute minimum (shown in Fig. 9b), and the value of the two identified unbalances are represented. Again the exciting forces are accurately identified.

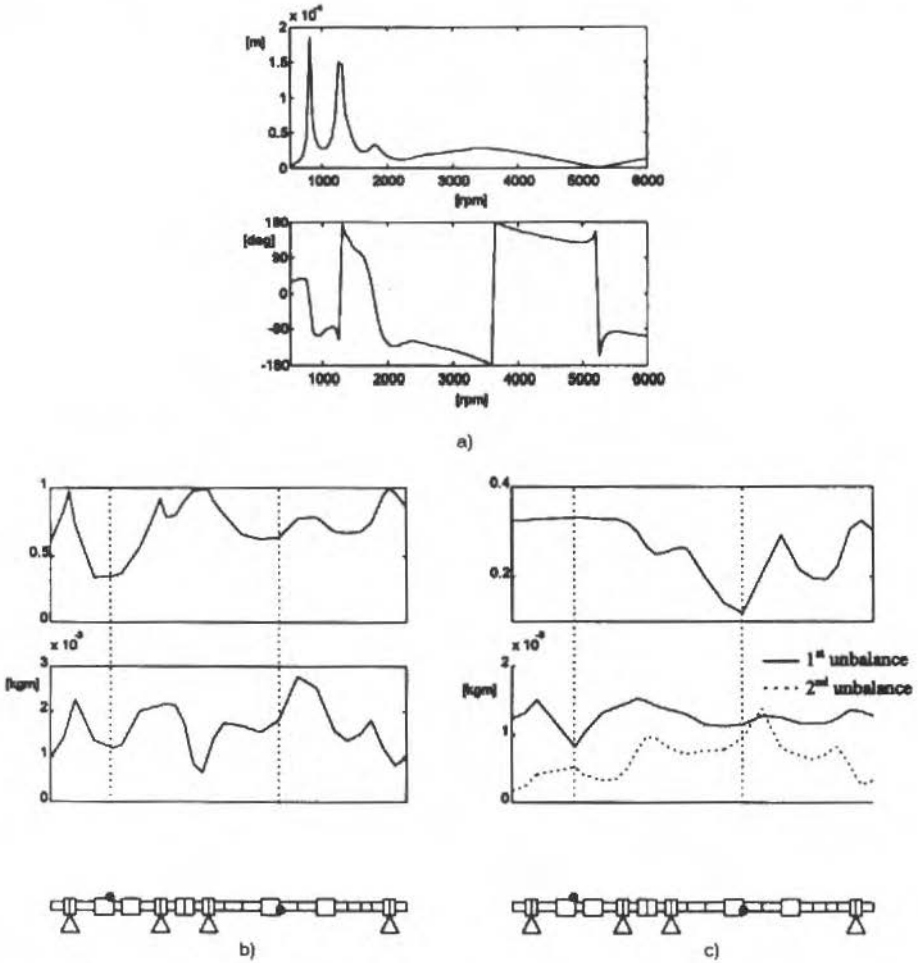


Fig. 9 Two Different Unbalances Placed on Disk 1 and 3 (0.001 kg each one): a) Frequency Response Curve at Bearing 1 (Journal's Vertical Displacement), b) Residuals With one Unbalance and Identified Unbalance Amount, c) Residuals With Two Unbalances and Identified Amounts.

Further in Fig. 10a the frequency response curve due to 2 unbalances placed on disks 1 and 2 (on the same shaft) and in Fig. 10b the corresponding residual are represented. In this case the 2 different unbalances cannot be identified because their positions are too close together.

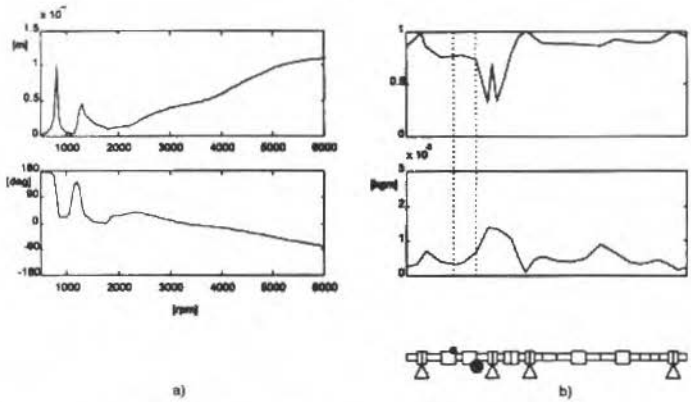


Fig. 10 Two Unbalances Placed on Disk 1 (0.001 kg ∠ 0°) and 2 (0.002 kg ∠ 180°): a) Frequency Response Curve at Bearing 1 (Journal's Vertical Displacement), b) Residuals With one Unbalance and Identified Unbalance Amount

As already said, different exciting causes can be modeled by a suitable system of forces: a concentrated unbalance is obviously represented by a rotating force, a concentrated bow (due to a coupling angular misalignment or to a rubbing in a seal) and a distributed bow (a thermal bow) can be represented by a system of equal and opposite couples (rotating with the shaft), which force the rotor to have a similar deformation shape. The problem is to see if, with the vibrations measured only in correspondence of the bearings, it is possible to distinguish between these different forcing systems. This problem has been faced in the paper of Bachschmid et al. (1995a) using simulated data and the complete linear model of the system: the results show a fairly good separation between different causes, when different rotating speeds are considered, although some exciting causes produce similar deformation shapes. A concentrated unbalance applied at the rotor span middle point and a distributed bow e.g. produce similar deformation shapes (and similar vibrations in the bearings) in the higher speed range, and the separation of the two different causes is possible only considering the low speed range in which the unbalance induced vibrations are much lower than the bow induced vibrations.

The identification of the cause of a bow is made by means of the location of the bow: if the bow is concentrated and located in correspondence of the coupling, the cause is probably a coupling misalignment, if it is located in correspondence of a (labyrinth) seal, the cause is probably a rub, if the bow is distributed along the central body of a generator or a steam turbine and is related to a change in operating conditions, a thermal bow may be supposed.

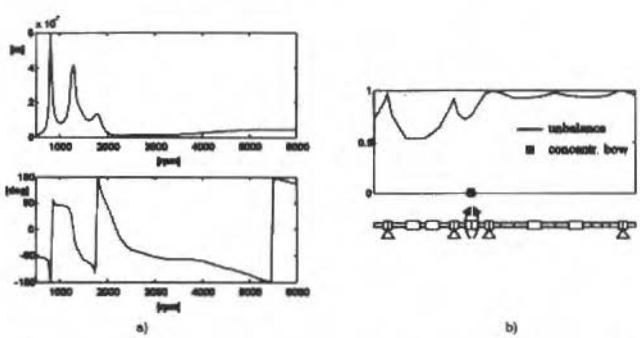


Fig. 11 Coupling Misalignment: a) Frequency Response at Bearing 1 (Journal's Vertical Displacement); b) Residuals Due to Unbalance and Concentrated Bow at Coupling Location

A comparison with the residuals obtained considering in each case all the different 1xrev. exciting causes is made in the following. Fig. 11a represents the frequency response curve generated by a misalignment coupling, and Fig. 11b represents the residual corresponding to an unbalance compared with the residual corresponding to the concentrated bow due to a coupling misalignment, in the position of the coupling between the two shafts. The lower minimum, close to zero, of the residual indicates that the exciting cause is the coupling misalignment, and also its values is accurately identified.

Finally Fig. 12a shows the frequency response curve due to a thermal bow in the central part of the second rotor (in between the two disks), and Fig. 12b the corresponding residuals (obtained weeping the rotor f.e. model with two couples, equal and opposite, spaced by two nodes, and plotting the corresponding residuals in correspondence of the middle point abscissa) compared with those generated by an unbalance: again the true exciting cause has been identified. Therefore the method seems to be suitable for identifying different 1xrev. exciting causes. It must be reminded that experimental data are affected by measuring errors, and rotor models and bearing models by modeling errors. An error sensitivity study shows that the identification method seems to be sufficiently robust with respect to all these different errors, so that its application to experimental results from power plants seems to be promising. This needs obviously accurate experimental validation, which will be performed by means of the test rig results, and by experimental results obtained on full size power plant machines.

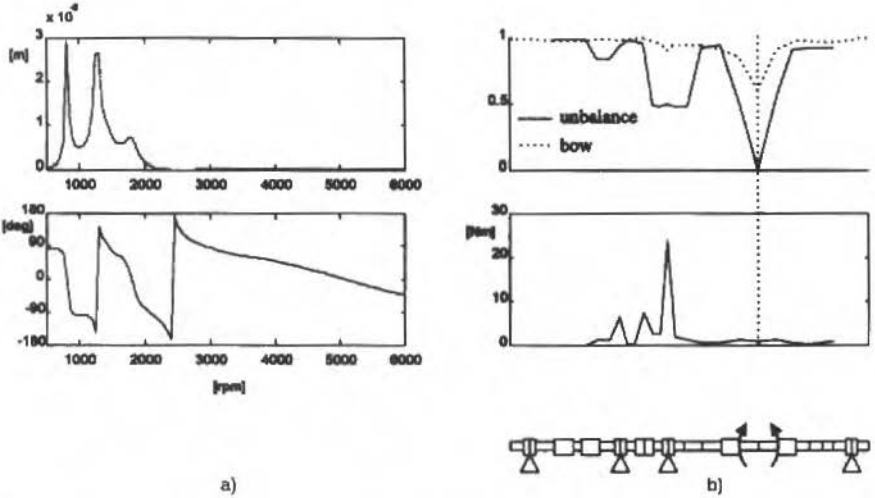


Fig. 12 Thermal Bow (Simulated by Two 1 Nm Bending Couples): a) Frequency Response at Bearing 1 (Journal's Vertical Displacement), b) Residual due to Unbalance and to Different Bowed parts of the Rotor, and Identified Values of Bending Moments

Conclusions

A method for identifying the exciting cause of the vibrations in rotor systems is presented, which, starting from relative and absolute shaft journal vibration measurements in the bearings and utilizing a linear f.e. model of the shaft and a non linear model of the oil film (and avoiding to represent with any model the supporting structure), allows to determine the type of exciting cause, its location and its amount and phase.

The method has been tested with numerical results, and seems to be able to distinguish between different exciting causes which produce similar effects, such as unbalance, bow, rub and coupling misalignments which all produce in certain conditions only 1xrev. vibration components.

The method will be tested also with experimental data.

Acknowledgments

This research is supported by the European Community (Brite Euram Contract BRPRCT95-0022).

References

Bachschmid, N., and Del Vescovo, D., 1990, "Thermal Bow in an Axi-symmetrical Shaft Containing a Transverse Notch and Related Vibrational Behaviour" 3.rd Int. Conf. on Rotordynamics Proc., Lyon, pp. 239-244.

Bachschmid, N., Bruni, S., and Collina, A., 1995a, "Sull'identificazione di Malfunzionamenti nei Rotori Tramite Misure di Vibrazioni nei Cuscinetti" AIMETA 95 Proc., Napoli, Vol. III, pp. 159-164 (in Italian).

Bachschmid, N., Bruni, S., and Collina, A., 1995b, "An Application of Unbalance Identification Techniques" IX IFToMM Congress Proc., Milano, pp. 1257-1261.

Bachschmid, N., Collina, A., and Bruni, S., 1995c, "The Effect of Misalignment in Rigid Couplings on the Statical and Dynamical Behaviour of Rotors" DINAME 95 Proc., Caxambu, pp. 201-204.

Bachschmid, N., Bruni, S., and Collina, A., 1996, "On the Identification of Rotor Bow, Coupling Misalignment and Unbalance in Rotor Systems from Bearing Measurements" COMADEM 96 Proceedings, Sheffield (UK), pp. 33-42.

Bachschmid, N., and Dellupi, R., 1997, "Non linear Behaviour of oil Film Bearings and its Relevance in Force Identification Procedures" Submitted to 1997 ASME Turbo Expo Congress to be held in Orlando (Florida).

Friswell, M.I., Lees, A.W., and Smaret, M.G., 1996, "Model Updating Techniques Applied to Turbogenerators Mounted on Flexible Foundations" II Int. Conf. in Structural Dynamic Modeling, Test, Analysis and Correlation, Cumbria.

Kanki, H., Yasuda, C., Umemura, S., Itoh, R., Miyamoto, C., and Kawaguchi, T., 1993, "Vibration Diagnostic Expert System for Steam Turbines" CISM-IFToMM Symposium on "Diagnostics of Rotating Machines in Power plants", Udine, Springer Verlag, pp. 25-36.

Krodkiewski, J.M., Zhang, N., and Ding, J., 1993, "Influence of Modeling of a Rotor System on Accuracy of Identification of its Unbalance Changes" CISM-IFToMM Symposium on "Diagnostics of Rotating Machines in Power Plants", Udine, Springer Verlag, pp. 291-302.

Lapini, G.L., Zippo, M., Bachschmid, N., Collina, A., and Vallini, A., 1993, "Experimental Tests and Model Based Calculations for the Diagnosis of a Crack in 320 MW Generator", CISM-IFToMM Symposium on "Diagnostics of Rotating Machines in Power Plants", Udine, Springer Verlag, pp. 87-98.

Tanaka, M., 1993, "The Diagnostic Technologies in Power Plants in Japan" CISM-IFToMM Symposium on "Diagnostics of Rotating Machines in Power Plants", Udine, Springer Verlag, pp. 199-210.

Vania, A., 1996, "Identification of the Modal Parameters of Rotating Machine Foundations" Internal Report, Dip. di Meccanica, Politecnico di Milano.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 176-191

ISSN 0100-7386
Printed in Brazil

A Simple State/Uncertainty Estimator for a Class of Uncertain Systems

Martin Corless
Purdue University
School of Aeronautics and Astronautics

Jay Tu
School of Industrial Engineering
West Lafayette, Indiana USA

Abstract

We consider here a class of uncertain systems consisting of a nominally linear part and an uncertain/nonlinear/time-varying part which can be regarded as a state-dependent/time-varying "disturbance input". Using only a measured output, we present simple estimators which can asymptotically estimate to any desired accuracy the system state and the disturbance input.
Keywords: *Mechanical Uncertain Systems, Nonlinear State Estimation, Nonlinear Filters.*

Introduction

We consider here a class of uncertain systems consisting of a nominally linear part and an uncertain/nonlinear/time-varying part which can be regarded as a state-dependent/time-varying "disturbance input". Using only a given measured output, we wish to asymptotically estimate the system state and the disturbance input. This problem is of practical significance because the model of any physical system contains uncertain terms due to uncertain inputs, parameters and nonlinearities.

The problem of designing an observer to estimate only the state of a system subject to unknown inputs has received considerable attention in the literature. In this existing literature, an observer design which assumes no a priori knowledge of the unknown input is called an unknown input observer (UIO). Basile et al (1969) and Guidorzi and Marro (1971) first investigated some structural aspects of the UIO using geometric concepts. Wang et al (1975) developed a trial and error procedure for constructing a reduced-order observer whose dynamics is completely decoupled from the unknown input. Bhattacharyya (1978) utilized geometric theory to obtain structural conditions which guarantee the existence of a reduced-order UIO. Kudva et al (1980) developed the now standard rank conditions for the existence of a reduced-order UIO. Several systematic procedures for designing reduced-order UIOs were proposed by dividing the state vector into two parts via a linear transformation; one part is directly driven by the unknown input and has to be measured completely, and the other part is estimated by the reduced-order UIO, which is decoupled from the disturbance. These procedures include the inversion algorithm of Kobayashi and Nakamizo (1982), the matrix generalized inverse method of Miller and Mukundan (1982), the matrix algebra method of Watanabe and Himmelblau (1982), the singular value decomposition techniques of Fairman et al (1984) and Park and Stein (1988), and the algebraic approaches of Breinl and Leitmann (1987), Guan and Saif (1991) and Hou and Müller (1992). Kureck (1982) considered a system with unknown inputs directly coupled to the output and proposed a full-order UIO. Yang and Wilde (1988) and Darouach et al (1994) developed alternative design procedures for full-order UIOs based on straightforward matrix calculations.

Compared with state estimation, very little research has been carried out on estimating unknown inputs. A common approach models the unknown input as the output of a linear system and incorporates the disturbance dynamics with the plant dynamics (Hostetter and Meditch, 1973; Johnson, 1976; Gourishankar et al, 1977; Müller, 1990). This approach is limited to specific types of unknown inputs. Park and Stein (1988) estimate unknown inputs by differentiating the output measurement. Their state and input observer is a combination of a reduced-order UIO and an algebraic equation relating the unknown input to the measured output and its derivative. Chen and Tomizuka (1988) and Chen (1990) considered a disturbance estimation problem for scalar systems. Tu and Stein (1995) proposed a model error compensator based on the output estimation error to

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

estimate the disturbance for an SISO system and incorporate it with an extended Kalman filter to estimate the state subject to the same disturbance. This approach was then extended to MIMO linear systems in Tu and Stein (1996).

Corless and Tu (in review) also consider the state estimation problem and provide equivalent characterizations of the UIO existence conditions of Kudva et al (1980). These characterizations yield insight into the design of UIO's and disturbance estimators which require differentiation of the output, e.g., the work of Park and Stein (1988). Using these equivalent characterizations, they then obtain an equivalent Lyapunov characterization of the UIO existence conditions. Provided the uncertain term and its rate of change satisfy some additional boundedness conditions, this Lyapunov characterization then permits the construction of a combined state/disturbance estimator. This estimator does not require differentiation of the measured output. Although exact asymptotic estimation is not achieved, one can asymptotically estimate the state and disturbance to any desired degree of accuracy.

In this paper, we consider the state/disturbance estimation problem for a more restrictive class of uncertain systems than that considered in Corless and Tu (in review). Here the number of measured outputs is the same as the number of "disturbance inputs". Because of the simpler structure of the systems under consideration, a very simple state/uncertainty estimator can be constructed. This estimator involves a single gain parameter γ. One can asymptotically estimate the state and disturbance to any desired degree of accuracy by choosing γ sufficiently large.

Problem Statement and Assumptions

We consider here uncertain systems described by

$$\dot{x} = Ax + Bf(t,x) \tag{1a}$$

where $x(t) \in R^n$ is the system state and $y(t) \in R^m$ is the measured output at time $t \in R$. The continuous function f, with $f(t,x) \in R^m$, models all uncertain/nonlinear/time-varying terms in the system description; we will regard this as an unknown state-dependent/time-varying disturbance input. The matrices A, B and C are known, constant, and of appropriate dimensions. Roughly speaking, we wish to construct an estimator which, using only the known system information and the measured output y, asymptotically estimates both the state x and the disturbance input f. To achieve this goal, we first introduce two assumptions.

Assumption 1 The matrix CB is symmetric and positive definite.

For SISO (scalar input scalar output) systems, this assumption is equivalent to the requirement that the transfer function given by

$$G(s) = C(sI - A)^{-1} B$$

has relative degree one and a positive "high frequency gain".

Assumption 2 For every complex number λ with non-negative real part,

$$rank\begin{bmatrix} A-\lambda I & B \\ C & 0 \end{bmatrix} = n+m \tag{2}$$

Remark 1 When the above rank condition fails for *any* complex number λ, there is a pair of vectors x_0, v with $x_0 \neq 0$ such that

$$(A - \lambda I)x_0 + Bv = 0$$

$$Cx_0 = 0$$

As a consequence of this, the solution of system (1) with initial condition $x(0) = x_0$ and $f(t,x) = e^{\lambda t}v$ satisfies

$$x(t) = e^{\lambda t}x_0$$

$$y(t) = 0$$

Hence, the initial state and input, x_0 and $e^{\lambda t}v$, are indistinguishable from zero initial state and input. Clearly, if the real part of λ is nonnegative, this is not desirable from the point of view of the above estimation desires. If λ has real part $-\alpha < 0$, then one cannot expect to achieve asymptotic estimation at an exponential rate greater than α.

Example 1 As an illustrative example, consider

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = f(t,x)$$

$$y = \alpha x_1 + x_2$$

where $\alpha > 0$. Here

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad C = \begin{bmatrix} \alpha & 1 \end{bmatrix}$$

Since $CB = 1$, assumption 1 holds. It can readily be shown that rank condition (2) holds except for $\lambda = -\alpha < 0$. Hence assumption 2 holds.

Some Preliminary Results

Before presenting estimators, we present some results which provide alternative characterizations of the first two assumptions. These results are used in demonstrating the properties of the proposed estimators and provide further insight into assumptions 1 and 2.

Assumption 1 and a Canonical Structure

Our first result, lemma 1, states that assumption 1 is equivalent to the existence of a state transformation

$$x = T\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}$$

so that the corresponding transformed system has the following structure:

$$\begin{aligned} \dot{\xi}_1 &= A_{11}\xi_1 + A_{12}\xi_2 + f(t,x) \\ \dot{\xi}_2 &= A_{21}\xi_1 + A_{22}\xi_2 \\ y &= C_1\xi_1 \end{aligned} \tag{3}$$

where C_1 is symmetric and positive definite.

Lemma 1 Suppose B and C are $n \times m$ and $m \times n$ matrices, respectively. Then the following statements are equivalent.

(a) There is a nonsingular matrix T such that

$$T^{-1}B=\begin{bmatrix} I_m \\ 0 \end{bmatrix}, \qquad CT=\begin{bmatrix} C_1 & 0 \end{bmatrix} \tag{4}$$

where C_1 is symmetric and positive definite.

(b) The matrix CB is symmetric and positive definite.

Proof:

(a) $\Rightarrow$ (b). If (a) holds, then

$$CB = CTT^{-1}B = C_1$$

Since C_1 is symmetric and positive definite, the same holds for CB.

(b) $\Rightarrow$ (a). We show that $R(B)$, the range of B, and $N(C)$, the null space of C, are complementary, i.e., every vector in R^n can be written as the unique sum of two vectors, one from each of these two spaces. First, we show that these two spaces only intersect at zero. Suppose x is common to $R(B)$ and $N(C)$. Then $x = Bu$ and $CBu = 0$. Since CB is positive definite, we must have $u = 0$ and, hence $x = 0$.

Since CB is positive definite, it is invertible and rank $CB = m$. From this it follows that

$$rank\ B = rank\ C = m$$

Hence the dimension do $R(B)$ is m and the dimension of $N(C)$ is $n - m$. Since the subspaces $R(B)$ and $N(C)$ only intersect at zero and the sum of their dimensions equals the dimension of R^n, these two spaces are complementary.

Hence if one chooses a matrix T_2 such that the columns of T_2 form a basis for the null space of C, then T_2 is $n \times (n - m)$ and the matrix

$$T:=\begin{bmatrix} B & T_2 \end{bmatrix}$$

is invertible.

From the definitions of the transformations, it follows that $T^{-1}B$ and CT have the structure as indicate in (4) where $C_1 = CB$. Also, C_1 is symmetric and positive definite.

Example 2 Recall example 1 and let $\xi_1 := \alpha x_1 + x_2$ and $\xi_2 := x_1$. Then, a transformed system with the canonical structure is given by

$$\begin{aligned} \dot{\xi}_1 &= \alpha\xi_1 - \alpha^2\xi_2 + f(t,x) \\ \dot{\xi}_2 &= \xi_1 - \alpha\xi_2 \\ y &= \xi_1 \end{aligned}$$

Assumption 2 and a Stability Condition

Suppose system (1) satisfies assumption 1 and recall the transformed system (3). In this section, we show that assumption 2 is equivalent to asymptotic stability of the subsystem

$$\dot{\xi}_2 = A_{22}\xi_2$$

First we have the following result.

Lemma 2 Suppose A, B, C are n x n, n x m and m x n matrices, respectively, and there exist nonsingular matrices T and C_1 such that

$$T^{-1}AT = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad T^{-1}B = \begin{bmatrix} I_m \\ 0 \end{bmatrix}, \quad CT = \begin{bmatrix} C_1 & 0 \end{bmatrix} \tag{5}$$

Then, a complex number λ is an eigenvalue of A_{22} if and only if

$$rank \begin{bmatrix} A - \lambda I & B \\ C & 0 \end{bmatrix} < n + m \tag{6}$$

Proof:

For any complex number λ, we have

$$\begin{bmatrix} T^{-1}AT - \lambda I & T^{-1}B \\ CT & 0 \end{bmatrix} = \begin{bmatrix} T^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} A - \lambda I & B \\ C & 0 \end{bmatrix} \begin{bmatrix} T & 0 \\ 0 & I \end{bmatrix}$$

Hence,

$$rank \begin{bmatrix} A - \lambda I & B \\ C & 0 \end{bmatrix} = rank \begin{bmatrix} T^{-1}AT - \lambda I & T^{-1}B \\ CT & 0 \end{bmatrix} = rank \begin{bmatrix} A_{11} - \lambda I & A_{12} & I \\ A_{21} & A_{22} - \lambda I & 0 \\ C_1 & 0 & 0 \end{bmatrix}$$

It now follows that

$$rank \begin{bmatrix} A - \lambda I & B \\ C & 0 \end{bmatrix} = rank \begin{bmatrix} 0 & 0 & I \\ A_{21} & A_{22} - \lambda I & 0 \\ C_1 & 0 & 0 \end{bmatrix} = rank \begin{bmatrix} A_{21} & A_{22} - \lambda I \\ C_1 & 0 \end{bmatrix} + m$$

Since C_1 is invertible, we have

$$rank \begin{bmatrix} A_{21} & A_{22} - \lambda I \\ C_1 & 0 \end{bmatrix} = rank \begin{bmatrix} 0 & A_{22} - \lambda I \\ C_1 & 0 \end{bmatrix} = rank\left(A_{22} - \lambda I\right) + m$$

Thus, inequality (6) holds if

$$rank\left(A_{22} - \lambda I\right) < n - m \tag{7}$$

Since A_{22} is an $(n - m)$ x $(n - m)$ matrix, the complex number λ is an eigenvalue of A_{22} iff rank condition (7) holds.

We have now the following corollary.

Corollary 1 Suppose A, B, C are $n \times n$, $n \times m$, and $m \times n$ matrices, respectively, and there exist nonsingular matrices T and C_1 such that (5) holds. Then, A_{22} is Hurwitz if rank condition (2) holds for every complex number λ with non-negative real part, i.e., assumption 2 holds.

Lyapunov-type Consequences of Assumptions 1 and 2

In this section, we show that assumptions 1 and 2 are equivalent to a Lyapunov-type condition, condition 1. Satisfaction of this condition yields a matrix P which is utilized in a Lyapunov analysis of the properties of the proposed estimators.

Condition 1 There exist a positive definite symmetric matrix P and a positive scalar μ such that

$$PA + A^T P - \mu C^T C < 0 \tag{8a}$$

$$B^T P = C \tag{8b}$$

We have now the main result of this section.

Lemma 3 Consider system (1) and assume B has full column rank. Then, condition 1 holds if assumptions 1 and 2 hold.

PROOF: Assumptions 1 and 2 imply condition 1. From lemma 1 it follows that, if assumption 1 holds then, there exists a transformation matrix T so that the transformed matrices

$$\hat{A} := T^{-1} A T, \qquad \hat{B} := T^{-1} B, \qquad \hat{C} := C T \tag{9}$$

have the structure given in (5) where and C_1 is positive definite. If we let

$$P = T^{-T} \hat{P} T^{-1} \tag{10}$$

where $\hat{P}$ is positive definite symmetric, then

$$\hat{P} = T^T P T \tag{11}$$

and requirements (8) on P are equivalent to the following requirements on $\hat{P}$

$$\hat{P}\hat{A} + \hat{A}^T \hat{P} - \mu \hat{C}^T \hat{C} < 0 \tag{12a}$$

$$\hat{B}^T \hat{P} = \hat{C} \tag{12b}$$

We now show that asymptotic stability of the matrix A_{22} implies the existence of a positive definite matrix P and a positive scalar μ satisfying (12) and, hence, using corollary 1, assumptions 1-2 imply condition 1. Since A_{22} is Hurwitz, there is a positive definite symmetric matrix P_{22} such that

$$-Q_{22} := P_{22} A_{22} + A_{22}^T P_{22} < 0$$

Letting

$$\hat{P} = \begin{bmatrix} C_1 & 0 \\ 0 & P_{22} \end{bmatrix},$$

it should be clear that $\hat{P}$ is positive definite symmetric. If we let

$$-\hat{Q} := \hat{P}\hat{A} + \hat{A}^T\hat{P} - \mu\hat{C}^T\hat{C}$$

then

$$\hat{Q} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{bmatrix}$$

with

$$-Q_{11} := C_1 A_{11} + A_{11}^T C_1 - \mu C_1^2$$
$$-Q_{12} := C_1 A_{12} + A_{21}^T P_{22}$$

Since $Q_{22} > 0$, we can use a Schur complement result and obtain that $\hat{Q} > 0$ (hence condition (12a) holds) if

$$Q_{11} > Q_{12} Q_{22}^{-1} Q_{12}^T$$

Letting

$$\tilde{Q}_{11} := C_1 A_{11} + A_{11}^T C_1 + (C_1 A_{12} + A_{21}^T P_{22}) Q_{22}^{-1} (A_{12}^T + P_{22} A_{21})$$

the above requirement is equivalent to

$$\mu C_1^2 > \tilde{Q}_{11}$$

or,

$$\mu > \lambda_{max}(C_1^{-1} \tilde{Q}_{11} C_1^{-1})$$

where, for a matrix M with real eigenvalues, $\lambda_{max}(M)$ denotes the largest eigenvalue of M. Finally, it can readily be seen that, as a consequence of the structure of $\hat{B}$, $\hat{C}$, and $\hat{P}$, one has $\hat{B}^T\hat{P} = \hat{C}$; i.e., condition (12b) holds.

Condition 1 implies assumptions 1 and 2. We first show that condition 1 implies that assumption 1 holds. Since P is positive definite and B has full column rank, the matrix B^TPB is positive definite symmetric. It follows from (8b) that $CB = B^TPB$. Thus CB is positive definite and symmetric.

It now follows from lemma 1 that there exists a transformation matrix T so that the transformed matrices defined in (9) have the structure given in (5) where C_1 is positive definite and symmetric.

We now show that condition 1 implies that A_{22} is Hurwitz and, hence, using corollary 1, condition 1 implies that assumption 2 holds.

Introducing the matrix $\hat{P}$ as defined by (11), $\hat{P}$ is positive definite symmetric and (as we have already shown) condition 1 implies that conditions (12) hold. Let

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \hat{P},$$

where P_{22} is an $(n - m) \times (n - m)$ and m is the number of columns of B. Equality (12b) implies that

$$P_{12} = 0$$

The corresponding "2 - 2" block of inequality (12a) implies that

$$P_{22}A_{22} + A_{22}^T P_{22} < 0 \tag{13}$$

Since P_{22} is symmetric and positive definite, this last inequality implies that A_{22} is a Hurwitz matrix.

A State/Disturbance Estimator

In this section we present a simple estimator for simultaneously estimating the state x and the uncertain term f of system (1). We first need some assumptions on f.

Assumptions on Uncertain Term

Assumption 3 There exist a known function f_0 and known non-negative constants β_1 and κ_1 such that

$$\left\| f(t,x) - f_0(t,\hat{x}) \right\| \le \beta_1 + \kappa_1 \left\| x - \hat{x} \right\|$$

for all $t \in R$ *and* $x,\ \hat{x} \in R^n$.

The function f_0 can be regarded as "nominal" f or an initial estimate of f; it could be zero.

Assumption 4 There exist known non-negative constants β_{22}, κ_{21}, κ_{22} such that along any solution $x(\cdot)$ of system (1),

$$\left\| \frac{df}{dt}(t,x(t)) - \frac{\partial f_0}{\partial t}(t,\hat{x}) - \frac{\partial f_0}{\partial x}(t,\hat{x})\dot{\hat{x}} \right\| \le \beta_2 + \kappa_{21} \left\| x(t) - \hat{x} \right\| + \kappa_{22} \left\| \dot{x}(t) - \dot{\hat{x}} \right\|$$

for all $t, \hat{x}, \dot{\hat{x}}$.

As a specific example of a term f satisfying the above two assumptions, consider an unknown, bounded, disturbance input w which has a bounded derivative, i.e.,

$$f(t,x)=w(t), \quad \|w(t)\| \le \beta_1, \quad \|\dot{w}(t)\| \le \beta_2$$

Here, assumptions 3 and 4 are assured with $f_0 = 0$. Next consider a term f due to known, bounded, time-varying parameters which have bounded derivatives, i.e.,

$$f(t,x)=F(t)x, \quad \|F(t)\| \le k_1, \quad \|\dot{F}(t)\| \le k_2$$

Considering $f_0(t,\hat{x})=F(t)\hat{x}$, one can readily show that assumptions 3 and 4 are assured with $\beta_1 = \beta_2 = 0$, $k_{21} = k_2$ and $k_{22} = k_1$. Finally f could be a combination of the above two types of terms, i.e.,

$$f(t,x)=F(t)x,+w(t)$$

State/Disturbance Estimator

The proposed state/disturbance estimators are described by

$$\boxed{\begin{aligned} \dot{\hat{x}} &= A\hat{x}+B\hat{f}, \qquad \hat{x}(0)=\hat{x}_0 \\ \hat{f} &= f_0(t,\hat{x})-\gamma(C\hat{x}-y) \end{aligned}} \tag{14}$$

where γ is a positive scalar, the initial estimate $\hat{x}_0$ of $x(0)$ is arbitrary, and $\hat{x}(t)$, $\hat{f}(t)$, are the estimates of $x(t)$, $f(t,x(t))$ respectively.

The next result, which is the main result of this paper, states that, by choosing γ sufficiently large, one can asymptotically estimate to any desired degree of accuracy the state and the uncertain term f of system (1).

Theorem 1 Consider system (1) subject to assumptions 1-2, estimator (14) and consider any ε_1, $\varepsilon_2 > 0$.

(a) If assumption 3 holds then, there is a $\gamma_1 \ge 0$ such that for all $\gamma \ge \gamma_1$ and all $\hat{x}_0$, we have

$$\limsup_{t\to\infty} \|\hat{x}(t)-x(t)\| \le \varepsilon_1$$

(b) If assumptions 3 and 4 hold then, there is a $\gamma_2 > 0$ such that for all $\gamma \ge \gamma_2$ and all $\hat{x}_0$, we have

$$\limsup_{t\to\infty} \|\hat{f}(t)-f(t,x(t))\| \le \varepsilon_2$$

Proof:

We first demonstrate (a). To this end, we introduce the state estimation error:

$\tilde{x} := \hat{x} - x$

Utilizing descriptions (1) and (14) of the plant and estimator, respectively, the evolution of the estimation error is described by

$$\dot{\tilde{x}} = (A - \gamma BC)\tilde{x} + B\left[f_0(t, x+\tilde{x}) - f(t,x)\right] \tag{15}$$

As a candidate Lyapunov function for the above error system, we consider

$$V(\tilde{x}) = \tilde{x}^T P \tilde{x}$$

where P along with some positive scalar $\mu > 0$ assure condition 1. Thus $C = B^T P$ and along any solution of the error system we have

$$\frac{dV(\tilde{x}(t))}{dt} = L(t, \tilde{x}(t))$$

where

$$\begin{aligned} L(t,\tilde{x}) &:= 2\tilde{x}^T PA\tilde{x} - 2\gamma\tilde{x}^T PBB^T P\tilde{x} + 2\tilde{x}^T PB\left[f_0(t,x(t)+\tilde{x}) - f(t,x(t))\right] \\ &\le \tilde{x}^T\left[PA + A^T P\right]\tilde{x} - 2\gamma\left\| B^T P\tilde{x}\right\|^2 + 2\left\| B^T P\tilde{x}\right\|\left[\kappa_1\left\|\tilde{x}\right\| + \beta_1\right] \\ &= -\tilde{x}^T Q\tilde{x} - (2\gamma - \mu)\left\| B^T P\tilde{x}\right\|^2 + 2\left\| B^T P\tilde{x}\right\|\left[\kappa_1\left\|\tilde{x}\right\| + \beta_1\right] \end{aligned}$$

where we have used assumption 3 and let

$$-Q := PA + A^T P - \mu C^T C < 0$$

For any two positive scalars μ_1 and μ_2, we have the following inequalities:

$$2\kappa_1\left\| B^T P\tilde{x}\right\|\left\|\tilde{x}\right\| \le \mu_1^{-1}\kappa_1^2\left\| B^T P\tilde{x}\right\|^2 + \mu_1\left\|\tilde{x}\right\|^2$$

$$2\beta_1\left\| B^T P\tilde{x}\right\| \le \mu_2^{-1}\beta_1^2\left\| B^T P\tilde{x}\right\|^2 + \mu_2$$

hence

$$L(t,\tilde{x}) \le -\tilde{x}^T\left[Q - \mu_1 I\right]\tilde{x} + \mu_2 - \left[2\gamma - \mu - \mu_1^{-1}\kappa_1^2 - \mu_2^{-1}\beta_1^2\right]\left\| B^T P\tilde{x}\right\|^2$$

If μ_1 is chosen so that

$$\mu_1 < \lambda_{min}(Q) \tag{16}$$

then $Q - \mu_1 I > 0$ and

$$\alpha := \lambda_{min}\left[P^{-1}(Q - \mu_1 I)\right]/2 > 0$$

If γ is chosen to satisfy

$$2\gamma \ge \mu + \mu_1^{-1}\kappa_1^2 + \mu_2^{-1}\beta_1^2 \tag{17}$$

then

$$L(t,\tilde{x}) \leq -2\alpha V(\tilde{x}) + \mu_2$$

From this one may deduce that

$$\|\tilde{x}(t)\| \leq r + c\|\tilde{x}(t_0)\|e^{-\alpha(t-t_0)}$$

for all $t \geq t_0$ where

$$c := \left(\lambda_{max}(P)/\lambda_{min}(P)\right)^{\frac{1}{2}}, \qquad r := c\left(\mu_2/2\alpha\right)^{\frac{1}{2}}$$

Considering any $\varepsilon_3 > 0$ and choosing any μ_2 so that

$$c\left(\mu_2/2\alpha\right)^{\frac{1}{2}} \leq \varepsilon_3 \tag{18}$$

yields

$$\|\tilde{x}(t)\| \leq \varepsilon_3 + c\|\tilde{x}(t_0)\|e^{-\alpha(t-t_0)} \quad \text{for all } t \geq t_0 \tag{19}$$

Choosing $\varepsilon_3 \leq \varepsilon_1$, we obtain the desired result.

We now demonstrate part (b). We first demonstrate that, under assumption 4, and any $\varepsilon_4 > 0$ there exists a $\gamma_3 > 0$ such that whenever $\gamma \geq \gamma_3$,

$$\limsup_{t\to\infty} \|\dot{\hat{x}}(t) - \dot{x}(t)\| \leq \varepsilon_4 \tag{20}$$

To this end, we let $z := \dot{\tilde{x}}$, and differentiate the state estimation error Eq. (15) to obtain

$$\dot{z} = (A - \gamma BC)z + Bg \tag{21}$$

where

$$g := \frac{d}{dt}(f_0(t,\hat{x}) - f(t,x))$$

This system has the same structure as that of the state estimation error equation (15). Since $\|\hat{x}(t) - x(t)\|$ is bounded, assumption 4 plays the role of assumption 3. Hence, using the same analysis as that used in the analysis of the estimation error dynamics, one can choose $\gamma_3 \geq 0$ such that for all $\gamma \geq \gamma_3$, one has

$$\|\dot{\tilde{x}}(t)\| \leq \varepsilon_4 + c\|\dot{\tilde{x}}(t_0)\|e^{-\alpha(t-t_0)} \qquad \text{for all } t \geq t_0 \tag{22}$$

where $\alpha > 0$; hence (20) holds.

Consider now the state estimation error Eq. (15) and rewrite it as

$$\dot{\tilde{x}} = A\tilde{x} + B\left[\hat{f} - f(t,x)\right]$$

Since the matrix B has full column rank, it has a left inverse B^L; thus $B^L B = I$ and

$$\hat{f} - f(t,x) = B^L\dot{\tilde{x}} - B^L A\tilde{x}$$

From this we obtain

$$\left\|\hat{f} - f(t,x)\right\| \le \left\|B^L\right\|\left\|\dot{\tilde{x}}\right\| + \left\|B^L A\right\|\left\|\tilde{x}\right\|$$

and hence

$$\limsup_{t\to\infty}\left\|\hat{f}(t) - f(t,x(t))\right\| \le \left\|B^L\right\|\varepsilon_4 + \left\|B^L A\right\|\varepsilon_3$$

By assuring that

$$\left\|B^L\right\|\varepsilon_4 + \left\|B^L A\right\|\varepsilon_3 \le \varepsilon_2$$

we obtain the desired result.

Application to a Mechanical System

Consider a mechanical system consisting of two small bodies, each of mass $m > 0$, connected via a linear spring of spring constant $k > 0$ and a dashpot of damping coefficient $c > 0$; see figure 1. The system is constrained to move along an inertially fixed horizontal line and its configuration can be completely described by the inertial displacements q_1, q_2 of the two bodies. The only available measurements are q_1 and $\dot{q}_1$. Suppose the system is subject to an unknown force $f(t)$ which acts on the first mass and we wish to estimate this force along with the position and velocity of the second mass. If f and its derivative $\dot{f}$ are bounded, we will show that we can apply the results of this paper to achieve the above estimation desires.

Applying Newton's second law to each mass yields

$$m\ddot{q}_1 + c(\dot{q}_1 - \dot{q}_2) + k(q_1 - q_2) = f$$
$$m\ddot{q}_2 - c(\dot{q}_1 - \dot{q}_2) - k(q_1 - q_2) = 0$$

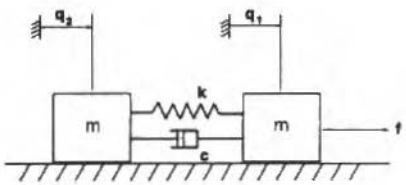


Fig. 1 A Mechanical System

If we introduce the states $x_1 = q_1$, $x_2 = q_2$, $x_3 = \dot{q}_1$, $x_4 = \dot{q}_2$ and the measured output $y = \alpha q_1 + \dot{q}_1$, where α is any positive number, this system has a state space description of the form (1) where

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{k}{m} & \frac{k}{m} & -\frac{c}{m} & \frac{c}{m} \\ \frac{k}{m} & -\frac{k}{m} & \frac{c}{m} & -\frac{c}{m} \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{m} \\ 0 \end{bmatrix}, \qquad C = \begin{bmatrix} \alpha & 0 & 1 & 0 \end{bmatrix}$$

Since $CB = 1/m$, assumption 1 holds. The rank condition (2) of assumption 2 fails when $\lambda = -\alpha$ or when λ is a root of the polynomial $m\lambda 2 + c\lambda + k$. Since this polynomial has all roots with negative real parts, assumption 2 holds. Hence, using only the measurement y and a simple estimator of the form (14), one can estimate, to any desired of accuracy, the disturbance f and the remaining states.

Numerical simulation results. For numerical simulation purposes, we considered

$$m = 2, \qquad k = 1, \qquad c = 3, \qquad \alpha = 1, \qquad f_0 = 0$$

In each simulation, the initial states of system and estimator were zero. The following two disturbances were considered.

(1) $f(t,x) = \sin(10\pi t) \cdot \sin(\pi t)$

(2) $f(t,x) = \sin((\sin(0.25\pi t) + 1)2\pi t)$

For each disturbance, two values of the gain parameter were chosen. The results are contained in Figs. 2-5.

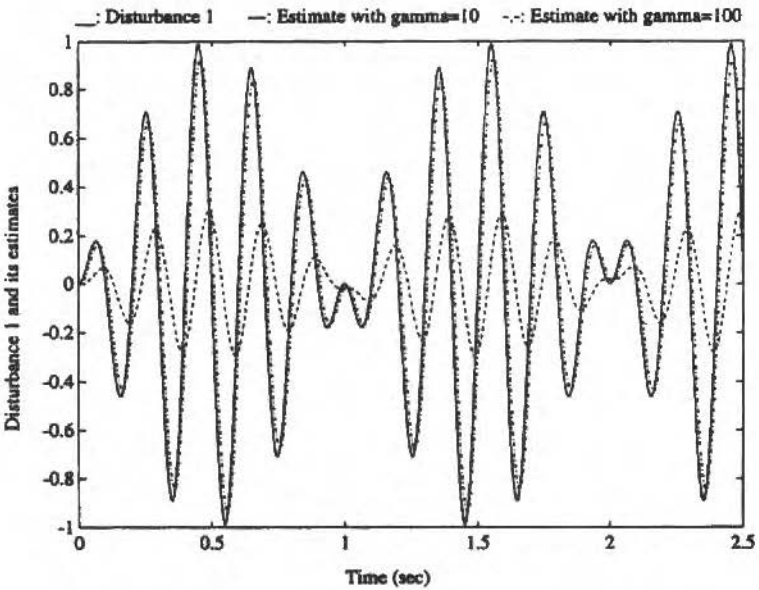


Fig. 2 Estimation of Disturbance 1

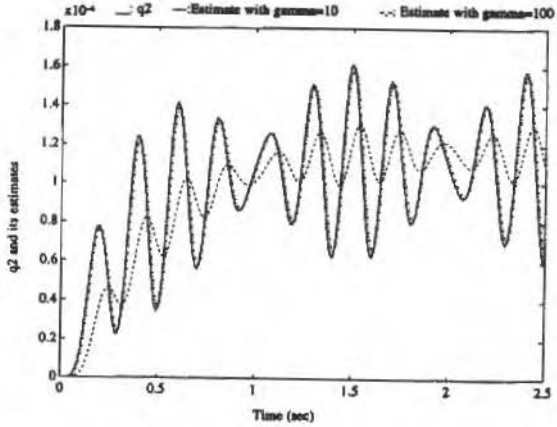


Fig. 3 State Estimation With Disturbance 1

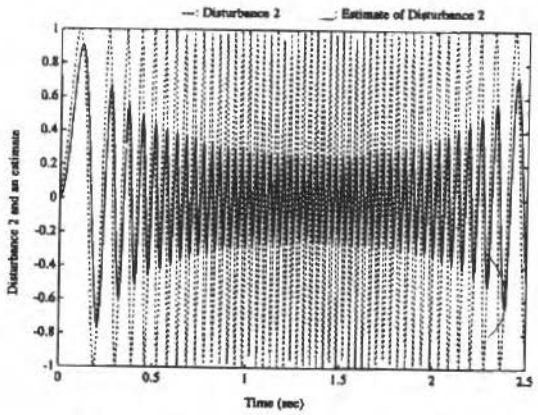


Fig. 4 Estimation of Disturbance 2 with $\gamma = 40$

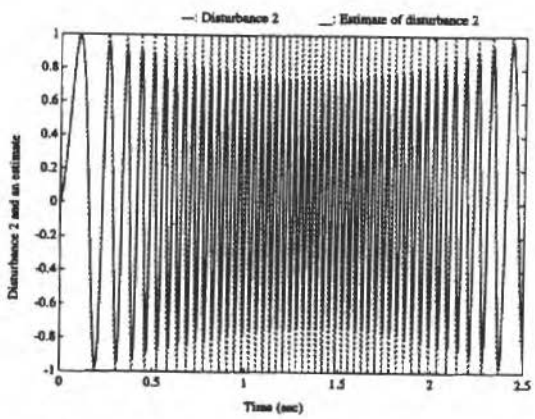


Fig. 5 Estimation of Disturbance 2 with $\gamma = 200$

Acknowledgments

This research was supported by the National Science Foundation under PYI grant MSS-90-57079 and RIA grant DMI 94-09985 and by the Purdue Engineering Research Center on Collaborative Manufacturing.

References

Basile, G., and Marro, G., 1969, "On the Observability of Linear, Time-Invariant Systems with Unknown Inputs", J. Optimiz. Theory Appl., Vol. 3, pp. 410-415.

Bhattacharyya, S.P., 1978, "Observer Design for Linear Systems with Unknown Inputs", IEEE Trans. Automat. Contr., Vol. AC-23, pp. 483-484.

Boyd, S. El Ghaoui, L., Feron, E., and Balakrishnan, V., 1994, "Linear Matrix Inequalities in System and Control Theory", SIAM.

Breinl, W., and Leitmann, G., 1987, "State Feedback for Uncertain Dynamical Systems", Applied Math. Comput., Vol. 22, pp. 65-87.

Chen, J., Patton, R.J., and Zhang, H-Y, 1996, "Design of Unknown Input Observers and Robust Fault Detection Filters", Int. J. Control, Vol. 63, No. 1, pp. 85-105.

Chen, M.S., 1990, "Uncertainty estimator", American Control Conference, pp. 2020-2024.

Chen, M.S., and Tomizuka, M., 1989, "Disturbance Estimator and its Application in Estimation of System Output Derivatives", Proceedings of the 28th IEEE Conf. Decision and Control, pp. 452-457.

Corless, M., and Tu, J.F., in review, "State and Disturbance Estimation for a Class of Uncertain Systems".

Darouach, M., Zasadzinski, M., and Xu, S.J., 1994, "Full-Order Observers for Linear Systems with Unknown Inputs", IEEE Trans. Automat. Contr., Vol. 39, pp. 606-609.

Fairman, F.W., Mahil, S.S., and Luk, L., 1984, "Disturbance Decoupled Observer Design via Singular Value Decomposition", IEEE Trans. Automat. Contr., Vol. AC-29, pp. 84-86.

Gahinet, P. Nemirovsky, A., Laub, A., and Chilali, M., 1995, "LMI Control Toolbox User's Guide", The Mathworks.

Guan, Y., and Saif, N., "A Novel Approach to the Design of Unknown Input Observers", IEEE Trans. Automat. Contr. Vol. AC-36, pp. 632-635.

Guidorzi, R., and Marro, G., 1971, "On Wonham Stabilizability Condition in the Synthesis of Observers for Unknown-Input Systems", IEEE Trans. Automat. Contr., Vol. AC-16, pp. 499-500.

Gourishankar, V., Kudva, P., and Ramar, K., 1977, Int. J. Contr. Vol. 25, pp. 311.

Hostetter, G., and Meditch, J.S., 1973, "Observing Systems with Unmeasurable Inputs", IEEE Trans. Automat. Contr., Vol. AC-18, pp. 307-308.

Hou, M., and Müller, P.C., 1992, "Design of Observers for Linear Systems with Unknown Inputs", IEEE Trans. Automat. Contr., Vol. AC-37, pp. 871-875.

Johnson, C.D., 1976, "Theory of Disturbance-Accommodating Controllers", In: Leondes, C.T. (Ed.)"Control and Dynamic Systems. Academic Press, Vol. 12.

Kobayashi, N., and Nakamizo, T., 1982, "An Observer Design for Linear Systems with Unknown Inputs", Int. J. Contr., Vol. 35, pp. 605-619.

Kudva, P., Viswanadham, N., and Ramakrishna, A., 1980, "Observers for Linear Systems with Unknown Inputs", IEEE Trans. Automat. Contr., Vol. AC-25, pp. 113-115.

Kurek, J., 1982, "Observation of the State Vector of Linear Multivariable Systems with Unknown Inputs", International Journal of Control, Vol. 36, pp. 511-515.

Miller, R.J., and Mukundan, R., 1982, "On Designing Reduced-Order Observers for Linear Time-Invariant Systems Subject to Unknown Inputs", Int. J. Contr., Vol. 35, pp. 183-188.

Müller, P.C., 1990, "Indirect Measurements of Nonlinear Effects by State Observers", IUTAM Sysmp. Nonlinear Dynamics in Engineering Systems, University of Stuttgart, Springer, Berlin, pp. 205-215.

Park, Y., and Stein, J.L., 1988, "Closed-Loop State and Input Observer for Systems with Unknown Inputs", International Journal of Control, Vol. 48, pp. 1121-1136.

Tu, J.F., 1991, "On-line Preload Monitoring for High-Speed Anti-Friction Spindle Bearings Using Robust State Observers", Ph.D. Dissertation, Department of Mechanical Engineering, The University of Michigan, Ann Arbor, MI.

Tu, J.F., and Stein, J.L., 1995, "On-line Preload Monitoring for High-Speed Anti-Friction Spindle Bearings", Journal of Dynamic Systems, Measurements, and Control, Vol. 117, pp. 43-53.

Tu, J.F., and Stein, J.L., 1996, "Modeling Error Compensation for Bearing Temperature and Preload Estimation", Journal of Dynamic Systems, Measurements, and Control, Vol. 118, pp. 580-585.

Wang, S.H., Davision, E.J., and Dorato, P., 1975, "Observing the States of Systems with Unmeasureable disturbances," IEEE Trans. Automat. Contr., Vol. 20, pp. 716-717.

Watanabe, K. and Himmelblau, D.M., 1982, "Instrument Fault Detection in Systems with Uncertainties", International Journal of System Science, Vol. 13, pp. 137-158.

Yang, F. and Wilde, R.W., 1988, "Observers for Linear Systems with Unknown Inputs", IEEE Trans. Automat. Contr., Vol. 33, pp. 677-681.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 192-206

ISSN 0100-7386
Printed in Brazil

Vehicle Modeling for Real Time Applications

G. Rill
FH Regensburg, Germany

Abstract

In the automotive industry enhanced control systems are more and more developed by using hardware or software in the loop techniques. For such applications an enhanced nonlinear vehicle model withreal time capacity is necessary. The paper presents a multi-purpose vehicle model where real time application was made possible not by simplifying the model but by using a special model technique, by adopting the generation of the equations of motion to the specific problems in vehicle dynamics, and by using a modified implicit Euler formalism for the numerical solution.
Keywords: *Vehicle Dynamics, Real-time Simulation.*

Introduction

The result of more then one hundred years of automotive engineering is a "high tech" product. To improve ride comfort and ride safety enhanced and "intelligent" spring- and damping elements with nonlinear and dynamic characteristics such as hydraulic bearings are used. In addition more and more active and semi-active systems were brought into action.

Thus, demands on vehicle modelling and vehicle simulation increased enormously.

Simple models separating longitudinal, lateral and vertical motions are no longer suitable for developping further enhancements. Due to severe nonlinearities in axle kinematics and force characteristics linearized models are too simple.

Models representing vehicle dynamics in good conformity to field test must satisfy the following demands:

- Nonlinear and threedimensional motion of the vehicle body;
- Nonlinear and if necessary even flexible wheel and axle kinematics;
- Detailed models of the steering system and drive train;
- Nonlinear and partly dynamic force element description, and
- Dynamic tire characteristics.

If the model should operate in driving simulator or in a hardware-in-the-loop test bench at least a

- Minimazed computer run time

is essential, (Rill, 1986).

In general computer codes derived by multi-purpose-algorithms (some are described in Schiehlen, 1990), are not adopted to specific problems in vehicle dynamics. The result is a computer code where model quality and run time performance are not optimized.

Increasing computer power in order to achieve real time application is not economic. Reducing model quality is an even poorer idea.

By tricky modeling techniques, by neglecting complicated and non relevant terms during mathematical description, and by using a modified Euler formalism for integrating the equations of motion real time applications with sophisticated vehicle models are possible even on small computers, (Rill, 1994).

On developping new control strategies or investigating the influence of certain design parameter to vehicle dynamics one benifits from a minimized computer run time too.

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

Vehicle Model

Structure

The vehicle is modeled by rigid bodies. In order to approximate the dynamic effect of a torsion flexible frame the vehicle body is devided into a front body and a rear body, Fig. 1.

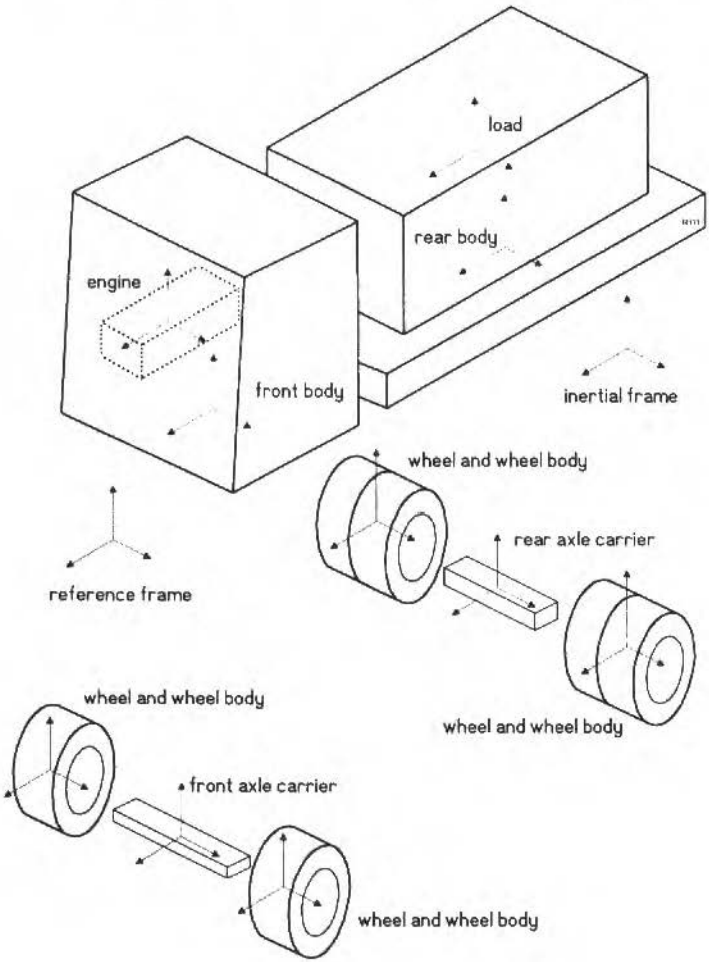


Fig. 1 Model Structure

Inertial frame 0 and reference frame B are fixed to road and respectively to front body.

At present the model is restricted to two axles. Every axle consists of an axle carrier, two wheel bodies, and two wheels. On independent axle suspension systems the axle carrier is omitted. Front wheel, rear wheel or all wheel steering is possible. Rigid rear axles may have single or double tired wheels.

The load is connected to the rear body. For sophisticated comfort analysis the engine suspension is modeled too.

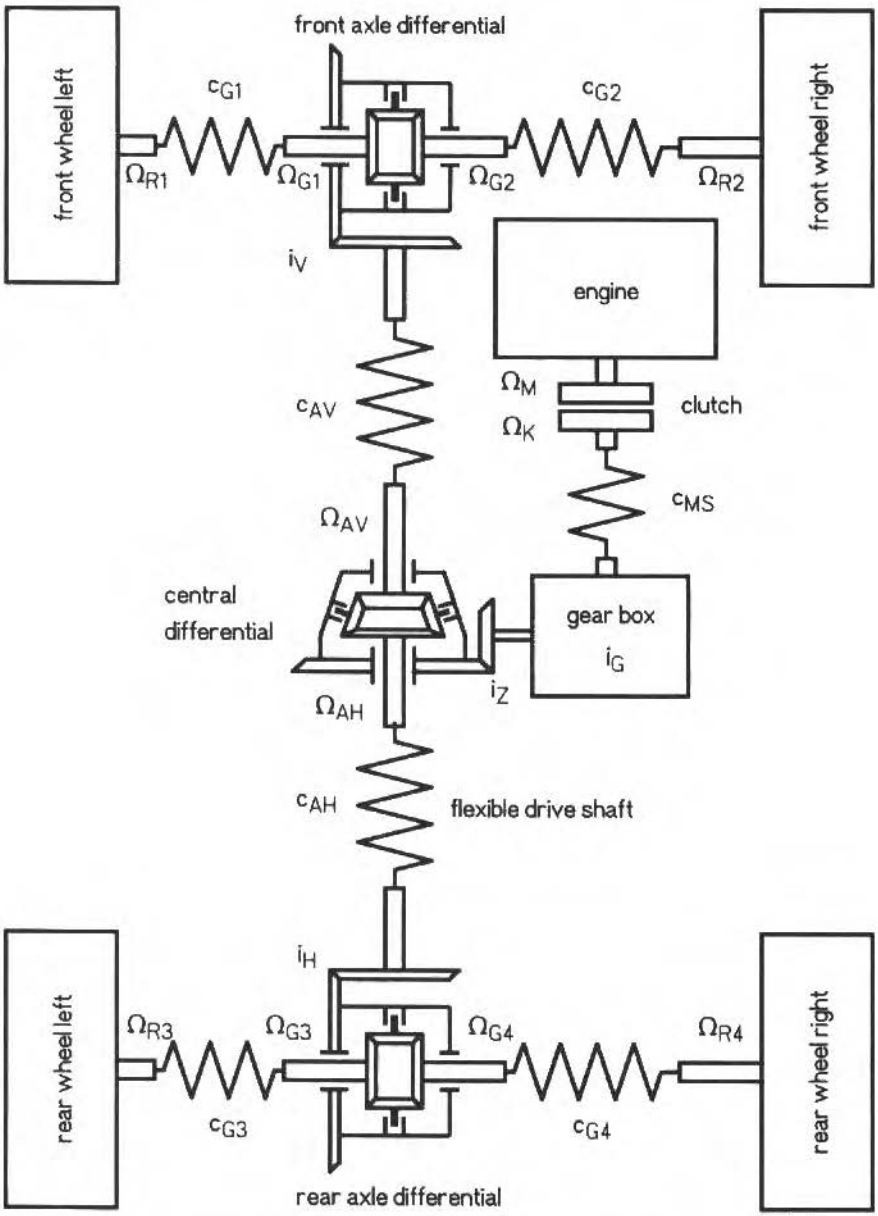


Fig. 2 Drive Train

A dynamic engine torque, nonlinear clutch characteristics, gear box, lockable differentials, and flexible shafts are the main components of the drive train, Fig. 2.

Modeling of braking torque includes stick slip effects.

The model for steering system is shown in Fig. 3. The steering model includes power steering.

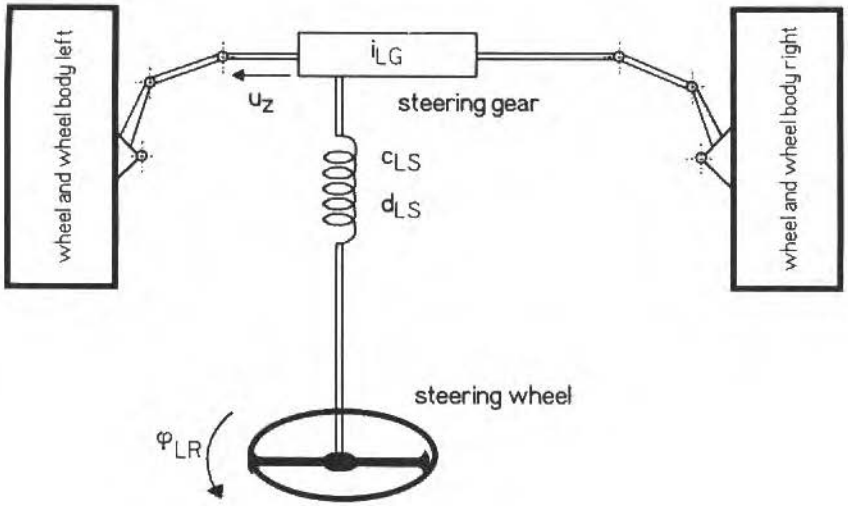


Fig. 3 Steering System

Different kinds of steering linkages are possible. The nonlinear steering kinematics is solved online.

The kinematics of wheel/axle suspension is calculated fully nonlinear. Pure kinematic or enhanced models with elastic bearings are possible, Fig. 4.

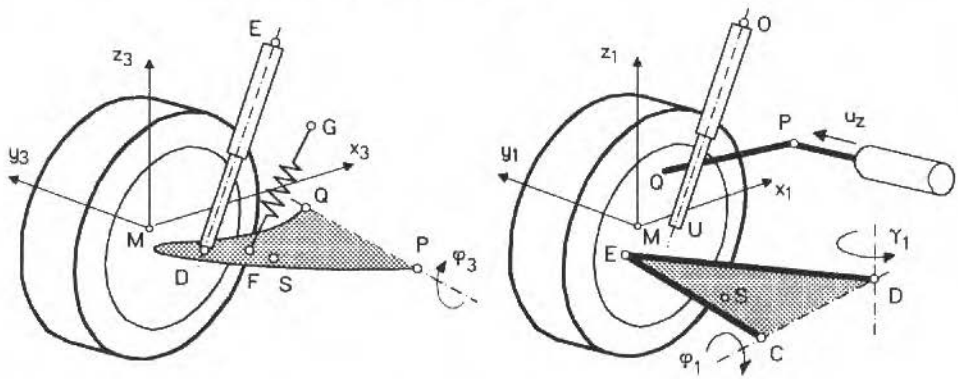


Fig. 4 Axle Kinematics

The suspension system may consist of conventional spring and damping elements, air springs, dynamic elements, or active and/or semi-active force elements.

Modeling the top mount makes it possible to include dry friction into the damping characteristics, Fig. 5.

Tire forces and torques are calculated by using a semi-physical approach. The "easy to use model" TMEASY includes contact geometry for arbitrary road profiles and a first order dynamic description of the longitudinal and lateral forces, Fig. 6.

Single obstacles or random road profiles are provided with the model. The coefficient of friction is part of the road model, thus μ-split situations can easily be simulated.

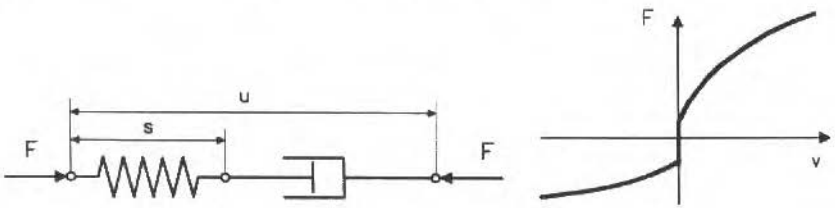


Fig. 5 Damper with Top Mount

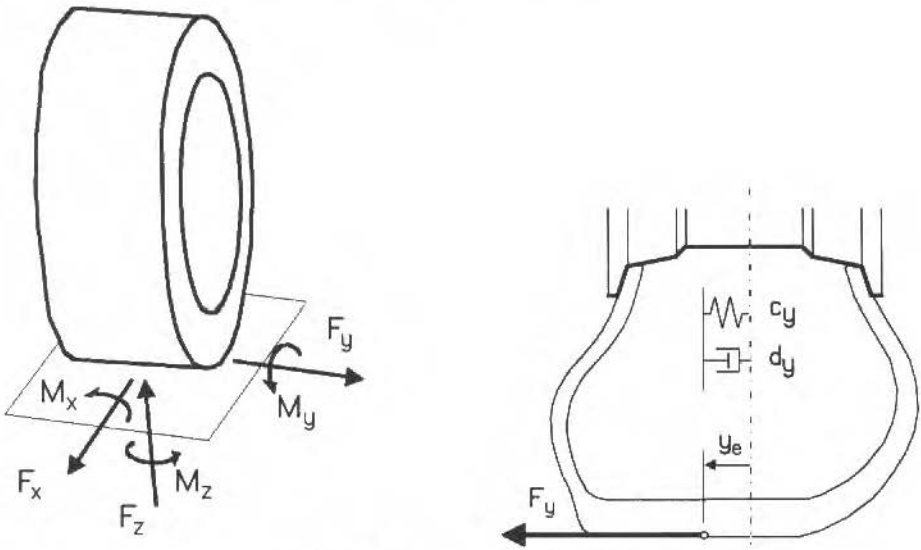


Fig. 6 Tire Forces and Torques

Model Variations

The model is not restricted to passenger cars. It can be easily adopted to different vehicle types, Fig. 7.

The passenger car model is characterized by a rigid car body and by independent axle suspension systems. In extension to the model described in Rill, 1994, it may have an elastically suspended engine.

Instead of an elastically mounted engine the truck model, c.f. Rill, 1986, has an elastically suspended driver's cab. Due to the flexible frame the rear body can perform a rotation around a longitudinal axis relative to the front body.

A rigid body, an independent front axle suspension, and a double tired rigid rear axle are typical features of modern coaches, (Reischl, 1995).

At agricultural tractors the rear axle is rigidly connected to the body. In most cases the front axle can only perform a rotation around a longitudinal axis relative to the body.

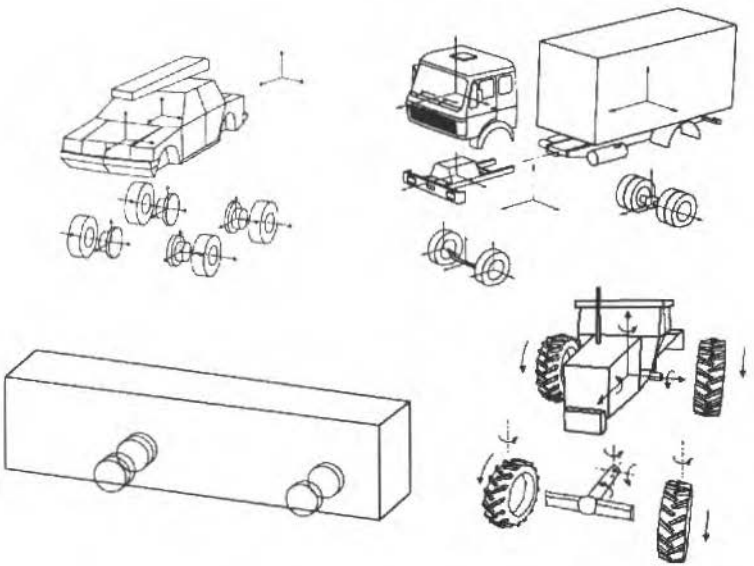


Fig. 7 Model Variations

Enhanced tractor layouts are equipped with a suspended driver's cab, an elastically suspended engine. In addition some tractors even have a suspended front axle, (Rill, Salg and Wilks, 1992).

A model extension to tank vehicles is described in Rill, 1995.

Kinematics

Reference Frame

The spatial motions of the vehicle are described by the momentary position and orientation of reference frame B with respect to the road-fixed inertial frame 0.

The position of frame B is given by the position vector

$$r_{0B,0} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{1}$$

Components of vector r_{0B} are denoted in the inertial frame 0. In (1) this is indicated by subscript 0 separated by a comma.

The orientation of the frame axis are described by a rotation matrix. Three elementary rotations are put together. The sequence

$$A_{0B} = \underset{yaw}{A_\gamma} \quad \underset{pitch}{A_\beta} \quad \underset{roll}{A_\alpha}$$

results in

$$A_{0B} = \begin{bmatrix} \cos\beta\cos\gamma & -\cos\alpha\sin\gamma + \sin\alpha\sin\beta\cos\gamma & \sin\alpha\sin\gamma + \cos\alpha\sin\beta\cos\gamma \\ \cos\beta\sin\gamma & \cos\alpha\cos\gamma + \sin\alpha\sin\beta\sin\gamma & -\sin\alpha\cos\gamma + \cos\alpha\sin\beta\sin\gamma \\ -\sin\beta & \sin\alpha\cos\beta & \cos\alpha\cos\beta \end{bmatrix} \tag{3}$$

Hence, the motion of the vehicle body is denoted by 6 generalized coordinates x, y, z, and α, β, γ. The velocity of the reference frame with respect to the inertial frame is given by

$$v_{0B,0} = \dot{r}_{0B,0} = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} \tag{4}$$

The velocity denoted in the inertial frame can be transformed to the reference frame

$$v_{0B,B} = A_{0B}^{T}\, \dot{r}_{0B,0} \tag{5}$$

In doing so the orthogonality of the rotation matrix

$$A_{B0} = A_{0B}^{-1} = A_{0B}^{T} \tag{6}$$

was already taken into consideration.

The angular velocity of the reference frame with respect to the inertial frame may be expressed directly in reference frame B

$$\omega_{0B,B} = \underbrace{\begin{bmatrix} 1 & 0 & -\sin\beta \\ 0 & \cos\alpha & \sin\alpha\cos\beta \\ 0 & -\sin\alpha & \cos\alpha\cos\beta \end{bmatrix}}_{K_B} \begin{bmatrix} \dot{\alpha} \\ \dot{\beta} \\ \dot{\gamma} \end{bmatrix} \tag{7}$$

The 6 components of $v_{0,B,B}$ and $\omega_{0,B,B}$ will now be chosen as generalized speeds.

First order kinematical differential equations connect generalized speeds with derivatives of generalized coordinates. From (5) and (7) one gets

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = A_{0B} \begin{bmatrix} v_{0Bx} \\ v_{0By} \\ v_{0Bz} \end{bmatrix} \tag{8}$$

and

$$\begin{bmatrix} \dot{\alpha} \\ \dot{\beta} \\ \dot{\gamma} \end{bmatrix} = K_B^{-1} \begin{bmatrix} \omega_{0Bx} \\ \omega_{0By} \\ \omega_{0Bz} \end{bmatrix} \tag{9}$$

Where the solution of (9) is given by

$$\dot{\gamma} = \left(\omega_{0Bz}\cos\alpha + \omega_{0By}\sin\alpha\right)/\cos\beta,$$
$$\dot{\beta} = -\omega_{0Bz}\sin\alpha + \omega_{0By}\cos\alpha, \qquad (10)$$
$$\dot{\alpha} = \omega_{0Bx} + \dot{\gamma}\cos\alpha.$$

Singularities occur for $\cos\beta \rightarrow 0$, respectively $\beta \rightarrow \pm 90°$. In normal driving conditions only the yaw angle γ reaches large values. Roll and pitch angles α, β are bounded.

The momentary state of the vehicle body, respectively the state of reference frame B is fully characterized by 6 generalized coordinates x, y, z, α, β, γ and 6 generalized speeds v_{0Bx}, v_{0By}, v_{0Bz}, ω_{0Bx}, ω_{0By}, ω_{0Bz}.

Relative Kinematics

The equations of motion are generated in the body-fixed reference frame B. The position and the orientation of frame B towards the inertial frame 0 is given by the position vector $r_{0B,0}$, (1) and the rotation matrix A_{0B}, (3).

Hence, position and orientation of body *i* towards the inertial can be expressed as

$$A_{0i} = A_{0B}\,A_{Bi} \qquad (11)$$

and

$$r_{0i,0} = r_{0B,0} + A_{0B}\,r_{Bi,B} \qquad (12)$$

The angular velocities $\omega_{0B,0}$ and $\omega_{Bi,B}$ can be derived from the rotation matrizes A_{0B} and A_{0i}, Rill, 1994. Hence, the angular velocity of body i with respect to the inertial frame is given by

$$\omega_{0i,0} = \omega_{0B,0} + A_{0B}\,\omega_{Bi,B} \qquad (13)$$

The velocity of body i with respect to the reference frame follows from (12)

$$\dot{r}_{0i,0} = \dot{r}_{0B,0} + \omega_{0B,0}\,x\left(A_{0B}\,r_{Bi,B}\right) + A_{0B}\,\dot{r}_{Bi,B} \qquad (14)$$

Transforming (13) and (14) to the reference frame one gets

$$\omega_{0i,B} = A_{0B}^{T}\,\omega_{0i,0} = \omega_{0B,B} + \omega_{Bi,B} \qquad (15)$$

and

$$\omega_{0i,B} = A_{0B}^{T}\,\dot{r}_{0i,0} = \underbrace{A_{0B}^{T}\,\dot{r}_{0B,0}}_{v_{0B,B}} + \omega_{0B,B}\,x\,r_{Bi,B} + \dot{r}_{Bi,B} \qquad (16)$$

Finally, the accelerations of body *i* towards the inertial frame 0 but denoted in reference frame B read as

$$\omega_{0i,B} = \dot{\omega}_{0B,B} + \omega_{Bi,B} + \omega_{0B,B}\,x\,\omega_{0i,B} \qquad (17)$$

and

$$a_{0i,B} = \dot{v}_{0B,B} + \dot{\omega}_{0B,B} x r_{Bi,B} + \ddot{r}_{Bi,B} + \omega_{0B,B} x \left(v_{0i,B} + \dot{r}_{Bi,B} \right) \tag{18}$$

If the state of reference frame B towards the inertial frame 0, given by A_{0B}, $r_{0B,0}$, $\omega_{0B,B}$, $v_{0B,B}$, and the derivatives $\dot{v}_{0B,B}, \dot{\omega}_{0B,B}$ are once determined only the relative quantities

$$A_{Bi}, r_{Bi,B}; \quad \omega_{Bi,B}, \dot{r}_{Bi,B}; \quad \dot{\omega}_{Bi,B}, \ddot{r}_{Bi,B} \tag{19}$$

are needed to calculate position, orientation, velocity, angular velocity and accelerations of each body towards the inertial frame.

In vehicle dynamics some complicated terms in $\dot{\omega}_{Bi,B}$ and $\ddot{r}_{Bi,B}$ can be neglected without loosing accurancy, Rill, 1994.

Dynamics

Jourdain's Principle

The equations of motion for the vehicle including steering system and drive train are generated using Jourdain's Principle. Given k rigid bodies it reads

$$\sum_{i=1}^{k} \left(\delta v_{0i,B}^T F_{i,B}^Z + \delta \omega_{0i,B}^T M_{i,B}^Z \right) = 0 \tag{20}$$

Separating the forces and torques applied to body i into constraint forces and torques $F_{i,B}^Z, M_{i,B}^Z$ and remaining terms $F_{i,B}^e, M_{i,B}^e$ the linear and angular momentum can be expressed as

$$\begin{aligned} m_i a_{0i,B} &= F_{i,B}^Z + F_{i,B}^e \\ T_{Si,B} \alpha_{0i,B} &= M_{i,B}^Z + M_{i,B}^e - \omega_{0i,B} x T_{Si,B} \omega_{0i,B} \end{aligned} \qquad i = 1(1)k \tag{21}$$

Where m_i denotes the mass of body i, and $T_{Si,B}$ is the tensor of inertia with respect to the center of gravity. The equations are denoted in reference frame B.

The virtual velocities $\delta v_{0i,B}$ and the virtual angular velocities $\delta \omega_{0i,B}$ are arbitrary infinitesimal velocities of the system compatible to the constraints.

The generalized coordinates and the generalized speeds necessary to define the state of the system are summarized in the vectors y and z. Then the virtual velocities and the virtual angular velocities can be expressed as

$$\begin{aligned} \delta v_{0i,B} &= \frac{\partial v_{0i,B}(y,z)}{\partial z_j}, \\ \delta \omega_{0i,B} &= \frac{\partial \omega_{0i,B}(y,z)}{\partial z_j}, \end{aligned} \qquad j = 1(1)n_z \tag{22}$$

where n_Z denotes the number of the generalized speeds.

Using (21) and (22), Eq. (20) reads as

$$\sum_{i=1}^{k}\left\{\frac{\partial v_{0i,B}^{T}}{\partial z_p}\left[m_i\,\frac{\partial v_{0i,B}}{\partial z_q}\dot{z}_q + m_i\,a_{0i,B}^{R} - F_{i,B}^{e}\right] + \frac{\partial \omega_{0i,B}^{T}}{\partial z_p}\left[T_{Si,B}\,\frac{\partial \omega_{0i,B}}{\partial z_q}\dot{z}_q + T_{Si,B}\,\alpha_{0i,B}^{R} + \omega_{0i,B}\times T_{Si,B}\,\omega_{0i,B} - M_{i,B}^{e}\right]\right\}\delta z_p = 0 \tag{23}$$

Acceleration terms not depending from derivatives of generalized speeds are collected in remaining accelerations

$$a_{0i,B}^{R} = \omega_{0B,B}\times v_{0i,B} + \frac{\partial v_{0i,B}(y,z)}{\partial y_j}\dot{y}_j,$$
$$\alpha_{0i,B}^{R} = \omega_{0B,B}\times \omega_{0i,B} + \frac{\partial \omega_{0i,B}(y,z)}{\partial y_j}\dot{y}_j, \tag{24}$$

Due to $\delta z_p \neq 0$ one gets from (23) the equations of motion

$$M(y)\dot{z} = Q(y,z) \tag{25}$$

with the mass matrix

$$M_{pq} = \sum_{i=1}^{k}\left\{\frac{\partial v_{0i,B}^{T}}{\partial z_p}\,m_i\,\frac{\partial v_{0i,B}}{\partial z_q} + \frac{\partial \omega_{0i,B}^{T}}{\partial z_p}\,T_{Si,B}\,\frac{\partial \omega_{0i,B}}{\partial z_q}\right\} \tag{26}$$

and the vector of the generalized forces and torques

$$Q_p = \sum_{i=1}^{k}\left\{\frac{\partial v_{0i,B}^{T}}{\partial z_p}\left[F_{i,B}^{e} - m_i\,a_{0i,B}^{R}\right] + \frac{\partial \omega_{0i,B}^{T}}{\partial z_p}\left[M_{i,B}^{e} - T_{Si,B}\,\alpha_{0i,B}^{R} - \omega_{0i,B}\times T_{Si,B}\,\omega_{0i,B}\right]\right\} \tag{27}$$

Two sets of first order differential equations, the kinematical differential equations

$$K(y)\dot{y} = z \tag{28}$$

defining generalized speeds, and (25), fully characterize the dynamics of a multi body system.

Structure of the equations of motion

The position vector of the vehicle is split into subvectors

$$y_G = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ x \\ y \\ z \end{bmatrix}, \; yv = \begin{bmatrix} y_{VA1} \\ \vdots \\ y_{VA_{nva}} \end{bmatrix}, \; yH = \begin{bmatrix} y_{HA1} \\ \vdots \\ y_{HA_{nva}} \end{bmatrix}, \; yL = [\delta], \; yA = [\alpha_2] \tag{29}$$

The angles α, β, γ and the coordinates x, y, z describe the position and orientation of the body fixed reference frame towards the inertial frame. The subsystems front and rear axle may have n_{VA} respectively n_{HA} degrees of freedom. The steering motion is described by the steering gear input angle δ, and the torsional deformation of the frame by the angle α_2.

The equations of motion can now be written down in the form

$$\underbrace{\begin{bmatrix} K_{GG} & 0 & 0 & 0 & 0 \\ 0 & E & 0 & 0 & 0 \\ 0 & 0 & E & 0 & 0 \\ 0 & 0 & 0 & E & 0 \\ 0 & 0 & 0 & 0 & E \end{bmatrix}}_{K} \underbrace{\begin{bmatrix} \dot{y}_G \\ y_V \\ y_H \\ y_L \\ y_A \end{bmatrix}}_{\dot{y}} = \underbrace{\begin{bmatrix} z_G \\ z_V \\ z_H \\ z_L \\ z_A \end{bmatrix}}_{z} \tag{30}$$

and

$$\underbrace{\begin{bmatrix} M_{GG} & M_{GV}^T & M_{GH}^T & M_{GL}^T & M_{GA}^T \\ M_{GV} & M_{VV} & 0 & M_{VL}^T & 0 \\ M_{GH} & 0 & M_{HH} & M_{HL}^T & M_{HA}^T \\ M_{GL} & M_{VL} & M_{HL} & M_{LL} & 0 \\ M_{GA} & 0 & M_{HA} & 0 & M_{AA} \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} \dot{z}_G \\ z_V \\ z_H \\ z_L \\ z_A \end{bmatrix}}_{\dot{z}} = \underbrace{\begin{bmatrix} Q_G \\ Q_V \\ Q_H \\ Q_L \\ Q_A \end{bmatrix}}_{Q} \tag{31}$$

where the partition used for the position vector y holds also for the speed vector z and the vector Q of the generalized forces and torques.

For the motion of the reference frame non-trivial generalized speeds were defined. In (30) this is expressed by $K_{GG}\,\dot{y}_G = z_G$. For the remaining coordinates trivial generalized speeds are used $\dot{y}_V = z_V, \dot{y}_H = z_H, \dot{y}_L = z_L, \dot{y}_A = z_A$.

Corresponding with the model structure the motions of the front and rear axle are kinematically uncoupled. In (31) this was already taken into consideration.

Numerical Solution

Implicit Euler Formula

If the implicit Euler formula is applied to (30), (31) it results in

$$K^{k+1}\left(y^{k+1}-y^{k}\right)=hz^{k+1},$$
$$M^{k+1}\left(z^{k+1}-z^{k}\right)=hQ^{k+1}. \tag{32}$$

where h is the step size. The superscripts k and $k+1$ denote quantities at time t and $t+h$.

A fully implicit solution is stable even for very large step sizes. But, in order to get a sufficient scanning of the road profile the step size must be limited anyway $h \le h_{max}$.

As the solution of (32) is very time consuming it is not a proper formula for real time applications.

For the integration of vehicle dynamics equations it is possible to derive from (32) a partial implicit Euler formula. This formula is characterized by minimized computer run time and sufficient stability conditions. Good experiences in the application of a partial implicit Euler formula were also reported in Hahn, 1991.

Partial Implicit Euler Formula

The equations of motion are written down in the body fixed reference frame. Due to the use of non-trivial generalized speeds the elements of the mass matrix are nearly constant

$$M^{k+1} \approx M^{k} \approx const \tag{33}$$

With the exception of submatrix K_{GG} all elements of the kinematical matrix K are constant. The elements of submatrix K_{GG} are depending from α, β, γ. These angles describe the orientation of the body fixed reference frame towards the inertial frame. Due to the inertia of the vehicle they will change from step to step only a little if the step size is not too large $\alpha^{k+1}-\alpha^{k} << 1, \beta^{k+1}-\beta^{k} << 1, \gamma^{k+1}-\gamma^{k} << 1$.

Thus, in good approximation it holds

$$K^{k+1} \approx K^{k}. \tag{34}$$

Then the first equation in (32) is simplified to

$$y^{k+1} = y^{k} + h\left(K^{k}\right)^{-1} z^{k+1} \tag{35}$$

The generalized forces and torques are depending from generalized coordinates and generalized speeds, Q = Q(y,z). The implicit expression

$$Q^{k+1} = Q\left(y^{k+1}, z^{k+1}\right) \tag{36}$$

is now approximated by

$$Q_G^{k+1} \approx Q_G^{k}, \tag{37}$$

$$\begin{aligned} Q_V^{k+1} \approx\ & Q_V\left(y_G^{k}, y_V^{k}+hz_V^{k}, y_H^{k}, y_L^{k}, y_A^{k}, z_G^{k}, z_V^{k}, z_H^{k}, z_L^{k}, z_A^{k}\right) \\ & + \frac{\partial Q_V}{\partial y_V}\left(y_V^{k+1}-\left(y_V^{k}+hz_V^{k}\right)\right)+\frac{\partial Q_V}{\partial z_V}\left(z_V^{k+1}-z_V^{k}\right) \end{aligned}, \tag{38}$$

$$Q_H^{k+1} \approx Q_H\left(y_G^k, y_V^k, y_H^k + hz_H^k, y_L^k, y_A^k, z_G^k, z_V^k, z_H^k, z_L^k, z_A^k\right) + \frac{\partial Q_H}{\partial y_H}\left(y_H^{k+1} - \left(y_H^k + hz_H^k\right)\right) + \frac{\partial Q_H}{\partial z_H}\left(z_H^{k+1} - z_H^k\right) \quad , \tag{39}$$

$$Q_L^{k+1} \approx Q_L\left(y_G^k, y_V^k, y_H^k, y_L^k + hz_L^k, y_A^k, z_G^k, z_V^k, z_H^k, z_L^k, z_A^k\right) + \frac{\partial Q_L}{\partial y_L}\left(y_L^{k+1} - \left(y_L^k + hz_L^k\right)\right) + \frac{\partial Q_L}{\partial z_L}\left(z_L^{k+1} - z_L^k\right) \quad , \tag{40}$$

$$Q_A^{k+1} \approx Q_A\left(y_G^k, y_V^k, y_H^k, y_L^k, y_A^k + hz_A^k, z_G^k, z_V^k, z_H^k, z_L^k, z_A^k\right) + \frac{\partial Q_A}{\partial y_A}\left(y_A^{k+1} - \left(y_A^k + hz_A^k\right)\right) + \frac{\partial Q_A}{\partial z_A}\left(z_A^{k+1} - z_A^k\right) \quad , \tag{41}$$

where only significant terms were taken into consideration.

With (30) and (34) the first equation in (32) is approximated by

$$\begin{aligned} y_G^{k+1} - y_G^k &= h\left(K_{GG}^k\right)^{-1} z_G^{k+1} \\ y_V^{k+1} - y_V^k &= hz_V^{k+1} \\ y_H^{k+1} - y_H^k &= hz_H^{k+1} \\ y_L^{k+1} - y_L^k &= hz_L^{k+1} \\ y_A^{k+1} - y_A^k &= hz_A^{k+1} \end{aligned} \tag{42}$$

and with (33), (37) to (41) the second equation in (32) results in

$$M\left(z^{k+1} - z^k\right) = h\left(Q^k + \frac{\partial Q}{\partial z}\left(z^{k+1} - z^k\right) + h\frac{\partial Q}{\partial y}\left(z^{k+1} - z^k\right)\right) \tag{43}$$

Introducing

$$M^{IE} = M - h\frac{\partial Q}{\partial z} - h^2\frac{\partial Q}{\partial y} \tag{44}$$

one gets

$$M^{IE}\left(z^{k+1} - z^k\right) = hQ^k \tag{45}$$

Following the approximations in (37) to (41) the partial derivatives are given by

$$\frac{\partial Q}{\partial z}=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\partial Q_V}{\partial z_V} & 0 & 0 & 0 \\ 0 & 0 & \frac{\partial Q_H}{\partial z_H} & 0 & 0 \\ 0 & 0 & 0 & \frac{\partial Q_L}{\partial z_L} & 0 \\ 0 & 0 & 0 & 0 & \frac{\partial Q_A}{\partial z_A} \end{bmatrix} \tag{46}$$

and

$$\frac{\partial Q}{\partial y}=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\partial Q_V}{\partial y_V} & 0 & 0 & 0 \\ 0 & 0 & \frac{\partial Q_H}{\partial y_H} & 0 & 0 \\ 0 & 0 & 0 & \frac{\partial Q_L}{\partial y_L} & 0 \\ 0 & 0 & 0 & 0 & \frac{\partial Q_A}{\partial y_A} \end{bmatrix} \tag{47}$$

Getting the new speeds

$$z^{k+1}=\left[z_G^{k+1}, z_V^{k+1}, z_H^{k+1}, z_L^{k+1}, z_A^{k+1}\right] \tag{48}$$

from (45) the new position subvectors $y_G^{k+1}, y_V^{k+1}, y_H^{k+1}, y_L^{k+1}, y_A^{k+1}$ follow from (42).

Applications

Driving Simulator

In a driving simulator real time capacity is essential.

The vehicle model is implemented on the Mercedes Benz Driving Simulator. It is used to develop enhanced suspension systems for coaches. The control strategy and the design parameters for a fully active and a semi-active suspension system are investigated, tested, and optimized.

In a very early design stage not only objective criterias such as wheel loads and body accelerations but also the subjective perceptions are available.

Different kinds of coach layouts can be studied. Environmental impacts, such as road roughness, slippy road, and wind gusts, can be taken into consideration.

By the complexity of the model it is granted that the results are in good conformity to real vehicle behaviour.

Just for a final test a prototype will be needed. Thus developping time and costs are reduced.

Hardware-in-the-Loop

The model was modified by the company TESIS to meet all requirements given by the German automotive company AUDI. For instance, a module describing the elastokinematics of modern axles was supplemented, and an interface to ADAMS kinematical calculations was provided.

AUDI wants a hardware-in-the-loop test bench to improve develop, and adopt modern slip control systems to new cars.

Field tests suffer from a lack of repeatability and are expensive and time consuming. Low friction tests are particularly difficult and limited to a few weeks in winter.

In a hardware-in-the-loop test road and vehicle are replaced by real time simulation, while the real electronic control unit and the real hydraulic systems are the same as in the vehicle.

Due to a very small step size which was needed to communicate with the hardware a multi processor system from dSPACE was used to achieve real time capacity.

References

Hahn, G. D., 1991, "A Modified Euler Method for Dynamic Analysis", International Journal for Numerical Methods in Engineering, Vol. 32, 943-955.

Reischl, M., 1995, "Untersuchung zum Einfluß Verschiedener Busparameter auf das Fahrverhalten von Bussen mit dem Simulationsprogramm MeFaS". Diplomarbeit, FH Regensburg, Fachbereich Maschinenbau.

Rill, G., 1986, "Fahrdynamik von Nutzfahrzeugen im Daimler-Benz Fahrsimulator". In: Berechnung im Automobilbau, VDI-Bericht 613, VDI-Verlag, Düsseldorf.

Rill, G., Salg, D., and Wilks, E., 1992, "Improvement of Dynamic Wheel Loads and Ride Quality of Heavy Agricultural Tractors by Suspending Front Axles", In: Heavy Vehicles and Roads, Ed.: Cebon, D. and Mitchell C.G.B., Thomas Telford, London.

Rill, G., 1994, "Simulation von Kraftfahrzeugen", Vieweg.

Rill, G., 1995, "Probleme mit Tankfahrzeugen". In: XII.International Heavy Vehicle Conference And EVU Annual Meeting, Budapest.

Schiehlen, W., 1990, "Multibody Systems Handbook", Springer, Berlin.

RBCM - J. of the Braz. Soc. Mechanical Sciences ISSN 0100-7386
Vol. XIX - No.2 - 1997 - pp. 207-216 Printed in Brazil

Weakly Nonlinear Second-order Dynamical Systems Identification Using a Random Parameters Linear Model

Christian Soize
O. Le Fur
Office National d'Etudes et de Recherches Aérospatiales
BP 72, 92322 Châtillon Cedex, France

Abstract

The objective of this paper is to present an identification procedure which is based on the use of a stochastic linearization method with random coefficients. The model is then defined as a multidimensional linear second-order dynamical system with random coefficients. An optimization procedure is developed to identify the parameters of the probability law of the random coefficients. The identification procedure is described step by step. Finally, an example is presented and shows the interest of the method proposed.
Keywords: *System Identification, Stochastic Linearization, Identification Algorithm.*

Introduction

For linear multidimensional second-order dynamical systems (m.s-o.d.s.) with time independent coefficients (constant coefficients), modal identification procedures are known and well developed. In this paper, we are interested in the identification of weakly nonlinear m.s-o.d.s. with constant coefficients, using a linear model and a stationary random input. The main idea is to use an infinite family of linear models to represent the nonlinear dynamical system i.e., a linear m.s-o.d.s. with random coefficients. This means that the weakly nonlinear m.s-o.d.s. is identified by a linear m.s-o.d.s. with uncertainties. Consequently, such an identification yields a linear model whose operator-valued frequency response function is a random stochastic process indexed by the frequency. The eigenfrequencies and associated eigenmodes are then deduced from the linearized representation which is identified (consequently, the eigenfrequencies are random variables).

Using broad-band stationary random excitation and constant coefficients of the model yields the classical Stochastic Linearization Method with Constant Coefficients (SLMCC). The SLMCC was introduced by Caughey in 1963 within the context of prediction methods. Many developments have been proposed in this area since this date and an excellent synopsis was made by Roberts and Spanos in 1990. An identification procedure based on SLMCC can be summarized as shown in Fig. 1 and will be refered in this paper as Method I.

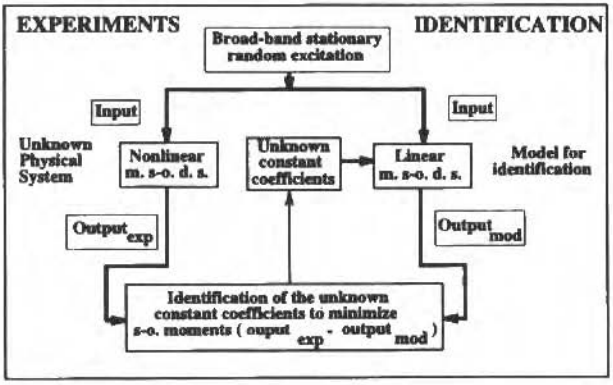


Fig. 1 Method 1 - Identification Procedure Based on a Stochastic Linearization Method With Constant Coefficients (SLMCC)

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

In this field of identification procedures, it should be noted that difficulties arise due to the presence of the mass matrix which is unknown and which must be identified. Fillâtre (1992) developed a method for identifying such an equivalent linear model in which mass, damping and stiffness matrices are constant and unknown. His approach is based on an extension of Kozin's works (1987, 1988) and can be considered as a method based on the SLMCC. Generally speaking, it is known that the SLMCC yields a very good approximation of the second-order (s-o.) statistical moments of the stationary response of second-order dynamical systems. Consequently, an identification method based on such a procedure yields an equivalent linear dynamical system which can restitute the second-order moments. Unfortunately, in some cases, although the second-order moments are correctly estimated, the matrix-valued spectral density function (s.d.f.) of the response may be erroneous. This difficulty was first shown by Miles for a one-dimensional nonlinear dynamical system Miles (1989). For such nonlinear dynamical systems, methods were proposed to calculate the power spectral density function of the stationary response without using Monte Carlo numerical simulation (Miles (1989), Bouc (1994) and Soize (1991, 1994a). The matrix-valued spectral density function can generally not be calculated explicitly for multidimensional nonlinear dynamical systems, except for particular cases related to linear dynamical systems with random parametric excitations (see for instance Soize (1994b). Recently, Bellizzi and Bouc (1995b) proposed an interesting method for multidimensional systems in the context of prediction methods.

The Stochastic Linearization Method with Random Coefficients (SLMRC) Soize (1991, 1994a) is adapted to identification procedures and allows the identification to be improved with respect to the classical SLMCC. This fact was recently proved by Soize (1995) for one-degree-of-freedom nonlinear second-order dynamical systems. This method, based on a linear dynamical model with random coefficients, has just been extended by Le Fur (1995) for the identification of weakly nonlinear multidimensional second-order dynamical systems and the details of the method can be found in Soize and Le Fur (1997). An identification procedure based on SLMRC can be summarized as shown in Fig. 2 and will be refered in this paper as Method 2. It should be noted that Method 2 uses Method 1. The purpose of this paper is to summarize a new approach developed by Soize and Le Fur (1997) for identifying weakly nonlinear multidimensional second-order dynamical systems based on the identification of a linear model with random coefficients.

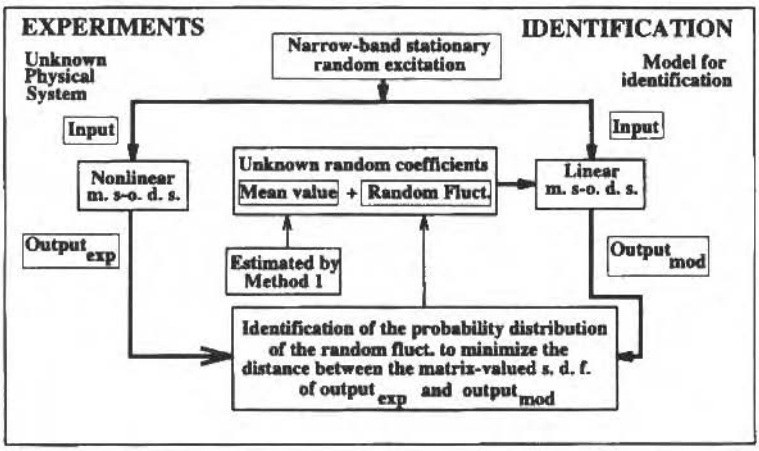


Fig. 2 Method 2 - Identification Procedure Based on a Stochastic Linearization Method With Random Coefficients (SLMRC)

Construction of the Model with Random Coefficients

We consider a weakly nonlinear dynamical system of dimension $n \geq 1$ subjected to an external random excitation. This dynamical system is written as the following stochastic differential equation

$$[M]\ddot{X}(t)+[C]\dot{X}(t)+[K]X(t)+\varepsilon \mathbf{f}(X(t))=\mathbf{F}(t) \tag{1}$$

in which matrices [*M*], [*C*] and [*K*] are positive definite; function **f** from $\mathbb{R}^n$ into $\mathbb{R}^n$ is odd, continuous and nonlinear; excitation force **F** is a Gaussian, second-order, centered, stationary, mean-square continuous stochastic process indexed by R with values in $\mathbb{R}^n$. The matrix-valued spectral density function of process **F** is written as

$$[S_{\mathbf{F}}(\omega)]=s(\omega)[B] \tag{2}$$

where [*B*] is a positive matrix and s is a positive-valued function defined on R having the required properties such that process **F** is physically realizable and approaches an ideal normalized narrow-band noise.

Within the context of an identification problem, it is assumed that stochastic differential equation (1) has a unique stationary, second-order, centered stochastic solution $\boldsymbol{X}$ having a matrix-valued spectral density function $[S_{\boldsymbol{X}}]$. Furthermore, function $[S_{\boldsymbol{X}}]$ is assumed to be square integrable on R.

Applying the identification procedure developed by Fillâtre (1992) (based on a stochastic linearization method with constant coefficients) yields the following linear stochastic differential equation on $\mathbb{R}^n$

$$[\underline{M}_c]\ddot{\mathbf{X}}(t)+[\underline{C}_c]\dot{\mathbf{X}}(t)+[\underline{K}_c]\mathbf{X}(t)=\mathbf{F}(t) \tag{3}$$

in which **F** is the stochastic process used in Eq. (1) and matrices $[\underline{M}_c]$, $[\underline{C}_c]$ and $[\underline{K}_c]$ result from the identification procedure and are positive definite.

We introduce the eigenmodes $\underline{\varphi} \in \mathbb{R}^n$ and the associated eigenfrequencies $\underline{\omega}$ of the conservative problem associated with Eq. (3), which are the solutions of the generalized eigenvalue problem $[\underline{K}_c]\underline{\varphi}=\underline{\omega}^2[\underline{M}_c]\underline{\varphi}$. Let $[\underline{\Phi}]$ be the (n x n) real matrix of the eigenmodes such that $[\underline{\Phi}]_{jk}=\{\underline{\varphi}_k\}_j$.

We introduce modal coordinates **Q** such that $\mathbf{X}=[\underline{\Phi}]\mathbf{Q}$. Substituting this change of coordinates in Eq. (3) yields

$$[\underline{M}_g]\ddot{\mathbf{Q}}(t)+[\underline{C}_g]\dot{\mathbf{Q}}(t)+[\underline{K}_g]\mathbf{Q}(t)=[\underline{\Phi}]^T\mathbf{F}(t) \tag{4}$$

where $[\underline{M}_g]=[\underline{\Phi}]^T[\underline{M}_c][\underline{\Phi}]$, $[\underline{C}_g]=[\underline{\Phi}]^T[\underline{C}_c][\underline{\Phi}]$ and $[\underline{K}_g]=[\underline{M}_g][\underline{\Omega}^2]=[\underline{\Phi}]^T[\underline{K}_c][\underline{\Phi}]$ are (n x n) real positive-definite matrices. Matrices and $[\underline{M}_g]$, $[\underline{K}_g]$ and $[\underline{\Omega}^2]$ are diagonal and $[\underline{C}_g]$ is a dense matrix in the general case.

We associate with Eq. (4) the following stochastic differential equation with random coefficients

$$[\underline{M}_g]\ddot{\mathbf{Y}}(t)+[\underline{C}_g]\dot{\mathbf{Y}}(t)+[\underline{K}_g]([I]+[\Lambda])\mathbf{Y}(t)=[\underline{\Phi}]^T\mathbf{F}(t) \tag{5}$$

where $[\Lambda]$ is a random variable with values in the (n x n) real diagonal matrices. We introduce the vector $\Lambda=(\Lambda_1,....,\Lambda_n)$ of its diagonal entries $\Lambda_i=[\Lambda]_{ii}$. It is assumed that $\{\Lambda_1,...,\Lambda_n\}$ are independent real-valued random variables. The probability law $P_{\Lambda_i}(d\lambda)$ of the real-valued random variable Λ_i is defined by a probability density function $p_{\Lambda_i}(\lambda)$ on R with respect to $d\lambda$:

$$P_{\Lambda_i}(d\lambda)=p_{\Lambda_i}(\lambda)d\lambda \tag{6}$$

in which for all $\lambda \in R$,

$$p_{\Lambda_i}(\lambda)=\alpha_{\Lambda_i}(1+\lambda)W_{\Lambda_i}(\lambda) \tag{7}$$

Real function $\lambda \mapsto W_{\Lambda_i}(\lambda)$ defined on R is such that

$$W_{\Lambda_i}(\lambda)=1_{\left[\lambda_i^{(1)},+\infty\right[}(\lambda)\left(\lambda-\lambda_i^{(1)}\right)e^{-\beta_{\Lambda_i}\left(\lambda-\lambda_i^{(1)}\right)^2} \tag{8}$$

Equations (6)-(8) define a parametric family of probabilities where the unknown parameters $\alpha_{\Lambda_i}, \beta_{\Lambda_i}$ and $\lambda_i^{(1)}$ verify the conditions $\alpha_{\Lambda_i}>0, \beta_{\Lambda_i}>0, 1+\lambda_i^{(1)}>0$. Since $P_{\Lambda_i}(d\lambda)$ is a probability, $P_{\Lambda_i}(R)=1$ and consequently, the three parameters $\alpha_{\Lambda_i}, \beta_{\Lambda_i}$ and $\lambda_i^{(1)}$ are dependent. Calculating α_{Λ_i} as a function of β_{Λ_i} and $\lambda_i^{(1)}$ yields

$$\alpha_{\Lambda_i}=\frac{2\beta_{\Lambda_i}}{1+\lambda_i^{(1)}+\frac{1}{2}\sqrt{\frac{\pi}{\beta_{\Lambda_i}}}} \tag{9}$$

Because of the independence of random variables $\{\Lambda_1,...,\Lambda_n\}$, the probability law of the R^n - valued random variable Λ is written as

$$P_\Lambda=\otimes_{i=1}^{n} P_{\Lambda_i} \tag{10}$$

It can be proved (Soize and Le Fur, 1997) that Eq. (5) has a unique second-order, centered, stationary solution $\mathbf{Y}$ which has a square integrable matrix-valued spectral density function given by the relation

$$[S_Y(\omega)]=\int_{R^n}[S_{Y_\Lambda}(\omega;\lambda)]P_\Lambda(d\lambda) \tag{11}$$

in which P_Λ is given by Eqs. (6)-(10) and matrix $[S_{\mathbf{Y}_\Lambda}(\omega;\lambda)]$ is such that

$$[S_{\mathbf{Y}_\Lambda}(\omega;\lambda)]=[H_\Lambda(\omega;\lambda)][\Phi]^T[S_{\mathbf{F}}(\omega)][\Phi][H_\Lambda(\omega;\lambda)]^* \tag{12}$$

$$\left[H_A(\omega;\lambda)\right]=\left[-\omega^2\left[\underline{M}_g\right]+i\omega\left[\underline{C}_g\right]+\left[\underline{K}_g\right]\left(\left[I\right]+\left[\lambda\right]\right)\right]^{-1} \tag{13}$$

Identification Procedure

The identification procedure consists in calculating the parameters of probability law P_A in order to minimize the "distance" between the matrix-valued spectral density function of the model responses and measured (experimental) responses.

Let **Q** be the R^n -valued stationary stochastic process such that $\boldsymbol{Q}=\left[\underline{\Phi}\right]^{-1}\boldsymbol{X}$ in which $\boldsymbol{X}$ is the measured stationary stochastic process (experimental responses) and $\left[\underline{\Phi}\right]$ are the estimated eigenmodes introduced in Section 2 . We then deduce that for all real ω, the matrix-valued spectral density function $\left[S_{\boldsymbol{Q}}(\omega)\right]$ of process $\boldsymbol{Q}$ can be written as

$$\left[S_{\boldsymbol{Q}}(\omega)\right]=\left[\underline{\Phi}\right]^{-1}\left[S_{\boldsymbol{X}}(\omega)\right]\left[\underline{\Phi}\right]^{-T} \tag{14}$$

Let $\lambda(1)$ and **A** be the vectors in R^n such that $\lambda^{(l)}=\left(\lambda_1^{(l)},\ldots,\lambda_n^{(l)}\right)$ and $\mathbf{A}=\left(A_1,\ldots,A_n\right)$ with

$$A_i=1/\sqrt{\beta_{A_i}},\ i\in\{1,\ldots,n\} \tag{15}$$

Let $\xi=\left(\xi_1,\ldots,\xi_n\right)$ be the vector in R^{2n} such that $\xi=\left(\mathbf{A},\lambda^{(l)}\right)$ and $\xi_i=\left(A_i,\lambda_i^{(l)}\right)$. Let D be the domain of ξ which is such that

$$D=\left\{\left(\mathbf{A},\lambda^{(l)}\right)\in R^{2n}\ \middle|\ A_i>0,1+\lambda_i^{(l)}>0,\forall i\in\{1,\ldots,n\}\right\} \tag{16}$$

In order to indicate the dependence of P_A and $[\mathbf{S_Y}]$ in ξ, we rewrite these quantities as P_A^{ξ} and $\left[S_{\mathbf{Y}}^{\xi}\right]$ respectively. Since the measured and the model matrix-valued spectral density functions are square integrable, the following cost function can be used,

$$H(\xi)=\sum_{i=1}^{n}\left\|\left[S_{\boldsymbol{Q}}\right]_{ii}-\left[S_{\mathbf{Y}}^{\xi}\right]_{ii}\right\|^2=\sum_{i=1}^{n}\int_R\left(\left[S_{\boldsymbol{Q}}(\omega)\right]_{ii}-\left[S_{\mathbf{Y}}^{\xi}(\omega)\right]_{ii}\right)^2 d\omega \tag{17}$$

The identification procedure is defined as the following optimization problem: find ξ_0 in D such that

$$H(\xi_0)=\min_{\xi\in D}H(\xi) \tag{18}$$

It should be noted that diagonal terms $\left[S_{\mathbf{Y}}^{\xi}\right]_{ii}$ depend on all the components of ξ due to the fact that matrix $\left[\underline{C}_g\right]$ used in the calculation of $\left[H_A(\omega;\lambda)\right]$ is not diagonal. In order to replace problem (18) by n independent optimization problems in R^2, elements $\left[S_{\mathbf{Y}}^{\xi}(\omega)\right]_{ii}$ are approximated by $\left[S_{\mathbf{Y}}^{\xi,i}(\omega)\right]_{ii}$ obtained by neglecting the extra-diagonal part in matrix $\left[\underline{C}_g\right]$. It should be noted that this approximation (introduced only to simplify the optimization problem) is not used in the final calculation of matrix $\left[S_{\mathbf{Y}}^{\xi}(\omega)\right]$ (See next section). Then, from Eqs. (2), (6)-(9) and (11)-(13), we deduce that for all i in $\{1, ..., n\}$

$$\left[S_{\mathbf{Y}}^{\xi,i}(\omega)\right]_{ii} = \frac{2e_{ii}s(\omega)}{1+\lambda_1^{(l)}+A_i\sqrt{\pi}/2}\int_0^{+\infty}\frac{\left(1+A_i x+\lambda_1^{(l)}\right)xe^{-x^2}\,dx}{\left(\left(1+A_i x+\lambda_1^{(l)}\right)\left[\underline{K}_g\right]_{ii}-\omega^2\left[\underline{M}_g\right]_{ii}\right)^2+\omega^2\left[\underline{C}_g\right]_{ii}^2} \tag{19}$$

in which $e_{ii}=\left\{\left[\underline{\Phi}\right]^T[B]\left[\underline{\Phi}\right]\right\}_{ii}$. For each i in $\{1,...,n\}$, we define the functional on $]0,+\infty[\times]-1,+\infty[\subset R^2$ such that

$$J_i(\xi_i)=\int_R\left\{\left[S_Q(\omega)\right]_{ii}-\left[S_{\mathbf{Y}}^{\xi,i}(\omega)\right]_{ii}\right\}^2 d\omega \tag{20}$$

Consequently, the optimization problem on a subset of R^{2n} defined by Eq. (18), is replaced by the n following optimization problems on a subset of R^2

$$J_i\left(\xi_{i,0}\right)=\min_{\xi_i\in]0,+\infty[\times]-1,+\infty[}J_i\left(\xi_i\right),\quad i\in\{1,...,n\} \tag{21}$$

It should be noted that each constraint optimization problem defined by Eq. (21) is not standard because J_i is not a convex function. Consequently the following method has been used:

Step1: Determine a bounded subdomain $C_i=\left[A_{2,i},A_{1,i}\right]\times\left[\lambda_{1,i}^{(l)},\lambda_{2,i}^{(l)}\right]$ included in unbounded domain $]0,+\infty[\times]-1,+\infty[$ such that C_i contains the solution (see [15]), in order to limit the space of the research for a solution.

Step 2: Use a global optimization algorithm on C_i based on an adaptive random search (Walter and Pronzato, 1994) which allows a first approximation $\xi_{i,a}$ of the solution of (21) to be constructed.

Step 3: Finally, use a local optimization method on C_i based on the Gauss-Newton algorithm and initialized with $\xi_{i,a}$, giving solution $\xi_{i,0}$ of problem (21).

Calculation of the Model Matrix-valued Spectral Density Function

Matrix-valued spectral densityfunction $\left[S_{\mathbf{Y}}(\omega)\right]$ can be calculated for $\xi = \xi_0$ where ξ_0 results from the identification procedure (See previous section). Knowing $\left[S_{\mathbf{Y}}(\omega)\right]$ which is the identified model of measured matrix $\left[S_Q(\omega)\right]$ expressed in terms of modal coordinates,we deduce the

identified model $[S_Z(\omega)]$ of the measured matrix $[S_X(\omega)]$ relative to the physical coordinates. We have the relation $[S_Z(\omega)]=[\underline{\Phi}][S_Y(\omega)][\underline{\Phi}]^T$ .Since matrix $[\underline{C}_g]$ is dense, the direct calculation of the model matrix-valued spectral density functiondefined by Eq. (11) requires calculating a n-uple integral on R^n for each ω.This calculation can only be carried out by numerical integration. Consequently, it cannot be made for large values of n (for instance when n is 10 or 20 (or more)).We therefore propose a construction of an approximation in Soize and Le Fur, (1997),which allows to calculate only simple integrals on R.

Example

For this example, an "experimental data base" is constructed using a Monte Carlo numerical simulation in the time domain of the second-order nonlinear dynamical system defined by Eq. (1) with n=5, where the nonlinear mapping **f** is defined by $\mathbf{f}(X(t))=\left([K]_{11}X_1(t)^3,\ldots,[K]_{55}X_5(t)^3\right)$ with ε =1875, where the frequency band of narrow-band process **F** is [14 Hz, 28 Hz] and where $[K]_{jj}$ are the diagonal terms of matrix $[K]$ appearingin Eq. (1). Matrices $[M]$, $[C]$ and [K] were generated by the formulas $[M]=[S]^{-T}[M][S]^{-1}$,$[C]=[S]^{-T}[C][S]^{-1}$ and $[K]=[S]^{-T}[k][S]^{-1}$, where

$$[S]=\begin{bmatrix} 0.208513 & 0.333334 & 0.301512 & 0.447214 & 0.447214 \\ 0.208513 & -0.333334 & 0.301512 & -0.447214 & 0.447214 \\ 0.625543 & 0.577350 & 0.522233 & 0.447214 & 0.0 \\ 0.625543 & 0.577350 & -0.522233 & -0.447214 & 0.0 \\ 0.361158 & -0.333334 & -0.522233 & 0.447214 & 0.774597 \end{bmatrix}$$

and where $[M]$ is the identity matrix (generalized masses equal to 1), $[K] = [M]\,[\Omega]^2$ in which $[\Omega]$ is the diagonal matrix whose diagonal is 2π x [18.0, 20.0, 20.4, 22.0 , 23.0] and finally, $[C]$ is a diagonal matrix whose diagonal is [4.5, 5.0, 5.13, 5.5, 5.8]. Digital signal processing on the time-simulated sample paths of the stationary response was applied to estimate the "measured" matrix-valued spectral density function $[S_X(\omega)]$ for ω in the frequency band of analysis.

It should be noted that some eigenfrequencies Ω_j of the underlying linear dynamical system associated with the nonlinear dynamical system are close (20.0 Hz and 20.4 Hz). In presence of nonlinearities, this kind of situation is generally recognized as a difficult problem within the context of structural dynamic identification. The procedure presented in Section 3 is used to identify the parameters ξ_j of the model. Figure 3 is related to the comparisons between the matrix-valued spectral density functions obtained by "experiments" and by identification of the model with constant coefficients. Figure 3 shows the comparison between $[S_X]_{ii}$ and $[S_X]_{ii}$ for i in {1,2,3,4,5}. These results correspond to those obtained by Fillâtre (1992). It should be noted that this first identification already yields a good identification (taking into account the intrinsic difficulties of the example

considered), but as mentioned in the first section, this kind of results can be improved using a more advanced model for identification (see below).

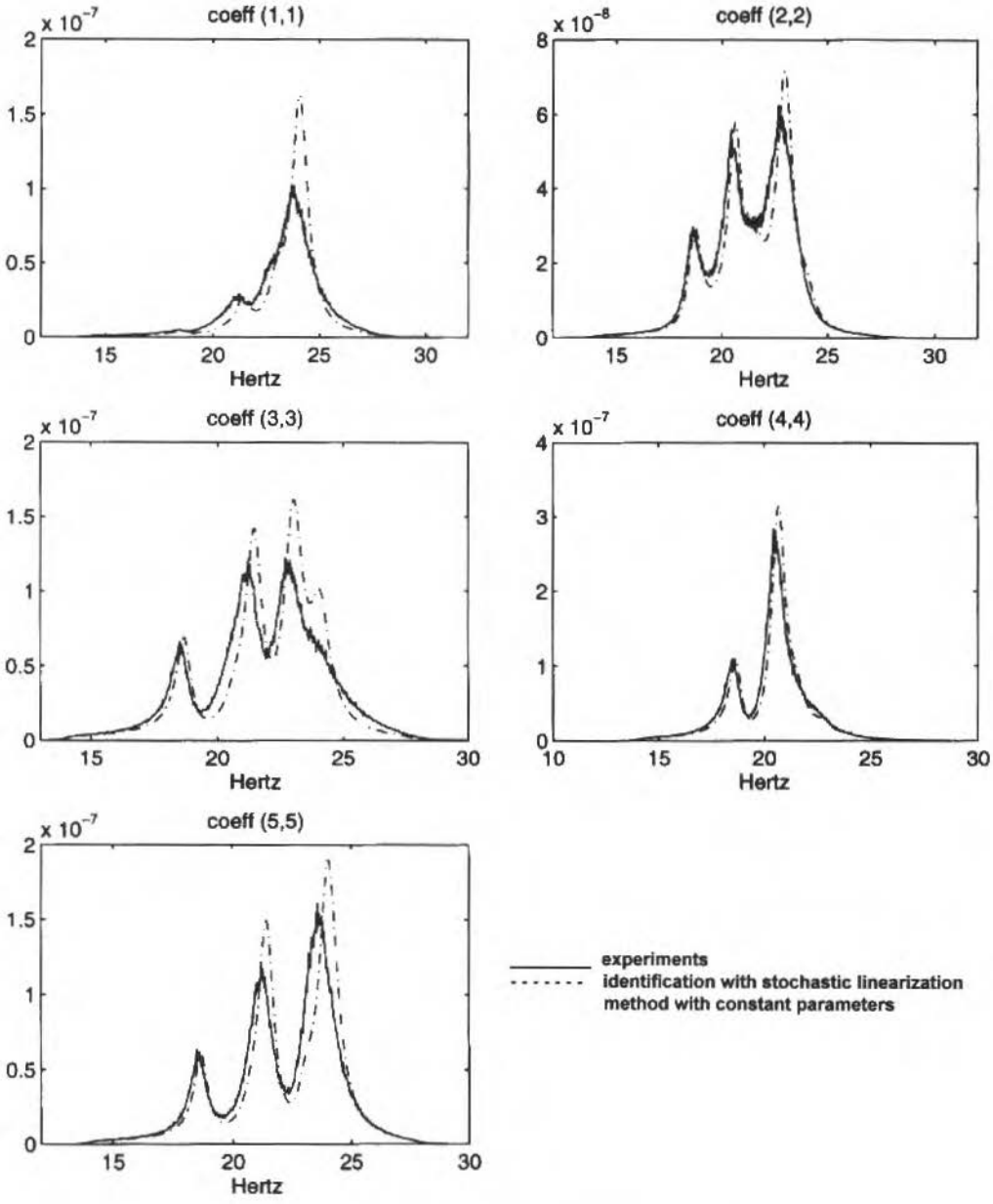


Fig. 3 Power Spectral Density Functions $[S_X]_{ii}$ and $[S_X]_{ii}$

Figure 4 is related to the comparisons between the matrix-valued spectral density functions obtained by "experiments" and by identification of the model with random coefficients for which the procedure was described in previous sections. Figure 4 shows the comparison between $[S_X]_{ii}$ and $[S_Z]_{ii}$ for i in {1,2,3,4,5}. It can be seen that the results obtained are much better than above. Complete results concerning this example are given in Soize and Le Fur (1997).

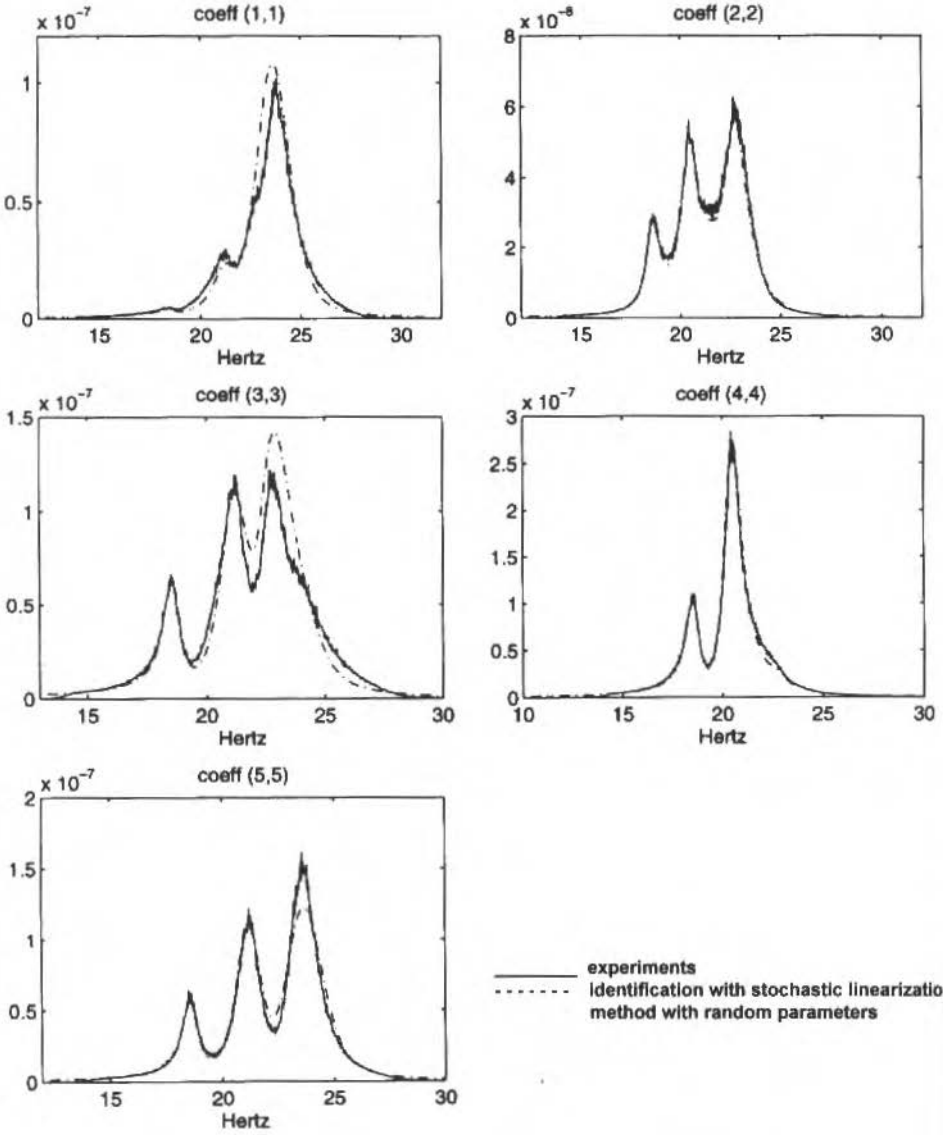


Fig. 4 Power Spectral Density Functions $[S_X]_{ii}$ and $[S_Z]_{ii}$:

Conclusion

This work started from previous research on modal identification of weakly nonlinear multidimensional second-order dynamical systems (based on the use of the equivalent stochastic linearization with constant coefficients). In some cases, this method has difficulty identifying the matrix-valued spectral density function of the stationary responses. In the present work, we have used the previous work to identify the mean part of the model. In order to improve the identification of spectral quantities, we developed a method based on stochastic linearization with random coefficients. This new identification procedure seems to be very efficient and can be implemented easily. The results show that this method yields better results than the previous one. Nevertheless, this method could be improved by introducing some statistical dependence between the components of the random coefficients expressed in the modal coordinates (or possibly by introducing extra-diagonal terms) in oder to model energetic exchanges between eigenmodes due to the weak nonlinearities. It should be noted that in this last case, the optimization problem introduced in the method could not be split into several optimization problems with a smaller size. This being the case, the efficiency of such a procedure would have to be studied with great care.

References

Bellizzi, S. and Bouc, R., 1995b, "2e Colloque National en Calcul des Structures", 16-19 mai 1995,Giens (France), Ed. Hermès, Tome 1, 75-80. Spectres de puissance de systèmes non-linéaires vibrants sous sollicitations aléatoires.

Bellizzi, S. and Bouc, R., 1995a, "Computational Stochastic Mechanics, P.D. Spanos, A.A., Balkema/Rotterdam/Brookfield", 77-86. Spectral response of asymmetrical random oscillators.

Bouc, R., 1994, "Journal of Sound and Vibration" 175(3), 317-331. The power spectral density of response for a strongly non-linear random oscillator.

Caughey, T.K., 1963 , "Journal of the Acoustical Society of America 35(11)", 1706-1711. Equivalent linearization techniques.

Fillatre, O., 1992 , "La Recherche Aérospatiale 3", 11-22. Identification of weakly nonlinear dynamic systems by means of random excitations.

Kozin, F., 1987, "IUTAM Symposium, Innsbruck". The method of statistical linearization for nonlinear stochastic vibrations. Nonlinear stochastic dynamic engineering systems.

Kozin, F., 1988, "Analysis and Estimation of Stochastic and Mechanical Systems W. Schiehlen, and W. Wedig (ed.)", 137-200. Structural Parameter Identification Techniques. New York: Springer-Verlag.

Le Fur, O., 1995, "Thèse de l'Université Paris 6". Identification modale des systèmes dynamiques multidimensionnels faiblement non linéaires par une méthode de linéarisation stochastique à paramètres aléatoires.

Miles, R.N., 1989, "Journal of Sound and Vibration 132(1)", 43-49. An approximate solution for the spectral response of Duffing's oscillator with random input.

Roberts,J.B. and Spanos, P.D. , 1990, "Random Vibration and Statistical Linearization". New York: Wiley.

Soize, C., 1994a, "Structural Safety and Reliability, G.I. Schueller, M. Shinozuka, J.T.P. Yao, A.A., Balkema/Rotterdam/Brookfield", 217-222. Stochastic linearization method with random parameters and power spectral density calculation.

Soize, C., 1991, "Rencontres Scientifiques du Cinquantenaire, Contrôle Actif Vibro-Acoustique et Dynamique Stochastique, Publications du L.M.A., C.N.R.S. No. 127", Sur le calcul des fonctions de densité spectrale des réponses stationnaires pour les systèmes dynamiques non linéaires.

Soize, C., 1994b, "The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit, Steady-State Solutions". Singapore : World Scientific.

Soize, C., 1995, "Probabilistic Engineering Mechanics". Stochastic Linearization Method With Random Parameters for SDOF Nonlinear Dynamical Systems: Prediction and Identification Procedures (Accepted).

Soize, C. and Le Fur, O., 1997, "Mechanical Systems and Signal Processing 11(1)". Modal identification of weakly nonlinear multidimensional dynamical systems using a stochastic linearization method with random coefficients.

Walter, E. and Pronzato, L., 1994 , "Identification de modèles paramétriques à partir de données, expérimentales". Paris: Masson.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 217-227

ISSN 0100-7386
Printed in Brazil

Tracking of Mechanical Systems Using Artificial Neural Networks

Henryk Flashner
T. Efrati
University of Southern California
Department of Mechanical Engineering
90089-1453 Los Angeles, CA USA
hflashne@alvitak.usc.edu

Abstract

A method for tracking control of mechanical systems using artificial neural networks is proposed. The proposed control law consists of a proportional and derivative control action, and a two-layer feedforward neural network used for on-line approximation of the nonlinear part of the system dynamics. Tuning of the neural network's weights is formulated in terms of a constrained optimization problem and solved on-line using a projection method. It is shown that the proposed control law yields closed-loop tracking error that tends asymptotically to zero while the control effort is minimized. The resulting algorithm has a simple structure and requires a very modest computation effort. The problem of tracking control for a two-degree of freedom planar manipulator is used to demonstrate the proposed method.
Keywords: *Control, Mechanical Systems, Neural Networks.*

Introduction

Feedback linearization is a basis for many approaches to trajectory tracking control of mechanical systems such as robots [Lewis et al. (1993); Spong and Vidyasagar (1989)]. Control laws designed by these methods consist of a two parts: a part that cancels the nonlinear terms in the plant dynamics rendering the nominal closed-loop system linear, and a part designed to satisfy stability and performance requirements. Exact feedback linearization assumes exact knowledge of the system model that is, in most cases, not available. To alleviate this shortcoming two classes of control design approaches were proposed. The robust control design approach that includes variable structure systems [Decarlo et al. (1988); Utkin (1978)], passivity based controllers [Lewis et al. (1993); Slotine and Li (1991)], Lyapunov-based control design [Chernousko (1996); Chernousko (1993); Corless (1993)]; and the adaptive control approach [Slotine and Li (1988)].

Neural networks controllers are viewed as belonging to the class of adaptive controllers. In neural network-based control laws, the role of the neural network is to generate a command that compensates for the uncertainties in the nonlinear dynamic model. The neural network operates in conjunction with a relatively simple controller, e.g. proportional+derivative, that is designed to ensure convergence of the tracking error. The idea is to exploit the nonlinear mapping ability of the neural network [Narendra and Parthasarathy (1991); Narendra and Parthasarathy (1990)] and the property of neural network as universal approximator [Cybenko (1989)], in order to identify the parameters needed for the computation of the feedback control law. Many existing neural networks-based controllers for mechanical systems suffer from important shortcomings as follows. From control design point of view the shortcomings include the fact that closed-loop performance is not guaranteed, and that the magnitude of the control effort is not part of the design. The shortcomings of the neural network algorithms include the fact that there is no systematic approach to initialization of the neural network's weights, that the neural network's weights are adjusted *off-line*, and the high complexity of the network configuration. Recent studies addressed some of these issues, see for example Lewis et al. (1995, 1996), and Sarangapani and Lewis (1990). However, the resulting neural network-based control laws possess complex structures and computationally intensive learning algorithms. Moreover, the important issue of limiting the magnitude of the control effort is not addressed in the derivation of the control law.

In this paper, a method for design of a neural network based approach to control mechanical systems that alleviates some of the shortcomings of existing algorithms is presented. The configuration of the proposed neural network and the tuning algorithm are relatively simple and therefore can be performed on-line resulting in a genuine learning feedback control law. It is shown

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

that the tracking error of the closed-loop system converges to zero is presence of uncertainties. Moreover the tracking is achieved while the control effort is minimized.

The paper is organized as follows. In Section "*Problem Formulation*", the problem of tracking control of mechanical systems is formulated. Fundamentals of neural networks is presented in Section "*Neural Networks*". The proposed control law is given in Section "*Neural Network Based Controller*". In Section "*Example...*" the performance of the proposed approach is demonstrated in a simulation example of a two-degree of freedom system. Concluding remarks are given in Section "*Conclusions*".

Problem Formulation

Dynamic Equations of Mechanical Systems

The equations of motion of an n-degrees of freedom mechanical system are given by:

$$M(q)\ddot{q}+V(q,\dot{q})+G(q)=\bar{u}+u+\tau_d \tag{1}$$

where $q(t) \in W \in R^n$ is a vector of generalized coordinates, W is a bounded region of the state space representing the workspace of the system, $\bar{u} \in R^n$ is a known reference input, $u \in R^n$ is a vector of generalized control forces, and $\tau_d \in R^n$ represents the disturbances acting on the system.

The dynamic equations given in Eq. (1) have the following characteristics (see [Lewis et al. (1993)]:

1. The mass matrix $M(q)$ is symmetric and positive definite for all q.

2. $M(q)$ is bounded from above and below as follows

$$0 \le \beta_{1q}^T q \le q^T M(q) q \le \beta_2 q^T q \qquad \beta_1, \beta_2 > 0 \tag{2}$$

3. The vector of Coriolis and centripetal terms $V(q,\dot{q})$ satisfies the following inequality

$$\left\| \bar{V}(q) \right\| \left\| \dot{q} \right\|^2 \le \upsilon_b \left\| \dot{q} \right\|^2 \quad \forall q \in W \tag{3}$$

4. It is always possible to find $\bar{V}(q,\dot{q})$ such that $V(q,\dot{q})=\bar{V}(q,\dot{q})\dot{q}$ and

$$S=\frac{1}{2}\dot{M}-\bar{V}(q,\dot{q}) \tag{4}$$

is skew symmetric.

5. The gravity term $G(q)$ is bounded as follows

$$\left\| G(q) \right\| \le g(q) \tag{5}$$

Control Design Approach

The objective is to design a control u such that the state $z=\begin{bmatrix} q & \dot{q} \end{bmatrix}^T$ of the system in Eq. (1), tracks the state of a model system $z_m=\begin{bmatrix} q_m & \dot{q}_m \end{bmatrix}^T$, where $q_m(t) \in R^n$ satisfies the equation

$$M_m(q_m)\ddot{q}_m + V_m(q_m,\dot{q}_m) + G_m(q_m) = \upsilon \tag{6}$$

The goal is to achieve satisfactory tracking in presence of system's parameter uncertainties, and of disturbances bounded as follows

$$\| \tau_d \| \leq \theta \tag{7}$$

The proposed control law is based on the inverse dynamics approach that consists of three parts

$$u = u_I + u_L - \ddot{u} \tag{8}$$

The term u_I is a linearizing control given by

$$u_I = M(q)\ddot{q}_m + V\dot{q} + G(q) \tag{9}$$

The term u_L consists of a linear Proportional + Derivative (PD) control logic

$$u_L = M(q)\left[K_d(\dot{q} - \dot{q}_m) - K_p(q - q_m)\right] \tag{10}$$

where $K_p, K_d \in R^{n \times n}$ are positive definite symmetric matrices. Assuming perfect knowledge of system's parameters and no disturbances, the closed-loop system is given by

$$\ddot{e}(t) + K_d\dot{e}(t) + K_d e(t) = 0 \tag{11}$$

where $e(t) = q(t) - q_m(t)$. Hence, in the ideal case, one can choose the PD controller parameters such that $z \rightarrow z_m$ arbitrarily fast. However, in presence of model uncertainties and disturbances, the linearizing controller u_I cannot eliminate completely the nonlinear terms. In this case, the closed-loop system is described by nonlinear, forced equations as follows

$$\ddot{e}(t) + K_d\dot{e}(t) + K_d e(t) = \chi(e(t),\dot{e}(t),t) \tag{12}$$

Obviously, asymptotic stability of these equations cannot be guaranteed.

The approach taken in this study, is to design a control law of the form of Eq. (8) with the linearizing u_I term being approximated by a neural network. The ultimate goal is to obtain a genuine feedback control system in which all computations are performed on-line.

Neural Networks

Notation and Definitions

Form an engineering stand point, a neural network can be viewed as a computing machine characterized by a parallel architecture, a similarity of neurons, and a set of adjustable weights. In this paper we consider N-layer feedforward neural network structure, in which the input and intermediate signals are always propagated forward, as shown in Fig. 1. A neural network is specified by its structure and a learning rule for the adjustable weights. We consider neural networks composed of neurons described by

$$y_i = \sigma(s_i) + u_i \qquad s_i = \sum_{j=1}^{n} w_{ij} x_j \tag{13}$$

where $\sigma(s) : R \to R$ is a nonconstant, bounded, and monotone increasing function of class C^k called an activation function. Typical selections for $\sigma(\cdot)$ is a sigmoid given by

$$\sigma(s) = \frac{1 - e^{-\alpha s}}{1 + e^{-\alpha s}} \tag{14}$$

For the network shown in Fig. 1, n_i (i = 1, ..., N) denotes the number of neurons in layer i, $y^0 \equiv x \in R^{n_0}$ is the input vector to the network (n_0 is the number of inputs to the network), $y^i \in R^{n_i}$ is the output of layer i, $\sigma^i \in R^{n_i}$ is defined in equation (14), $y^N \in R^{nN}$ is the output of the network, $W^i \in R^{n_i x n_{i-1}}$ is the weight matrix between layer i and layer i - 1, and $u^i \in R^{n_i}$ is the bias vector of layer i. The output of each layer is given by

$$y^i = \sigma^i(W^i y^{i-1}) + u^i \qquad i = 1, \ldots, N \tag{15}$$

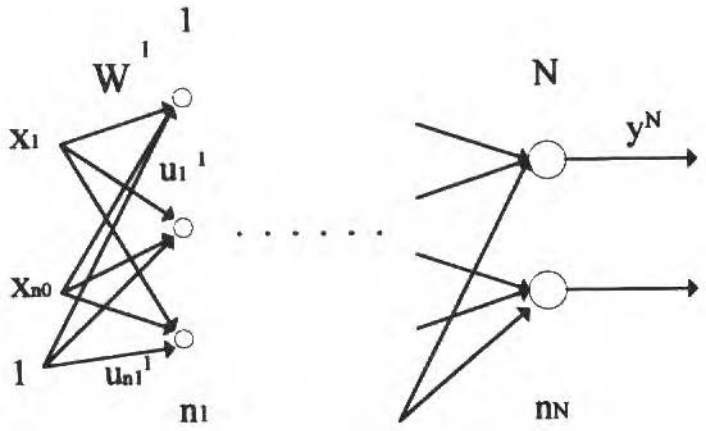


Fig. 1 A Feedforward Neural Network

Neural Network as Universal Approximator

Consider a two-layer neural network given by

$$y^2 = W^2 \sigma(W^1 x + u^1) + u^2 \tag{16}$$

The following theorem states that a two-layer feedforward neural network (N.N), with a sufficient number of hidden units, with an activation function given in (14), and linear output units, is capable of approximating any continuous function $f: R^n \to R^m$, to any desired accuracy.

Theorem 1 [Cybenko (1989)] Let σ be any continuous activation function. Then given any continuous real valued function f(·), on a compact subset $s \subset R^n$ and $\varepsilon > 0$, there exist vectors w_1, ..., w_n, α and θ and a parametrized function $G(\cdot, w, \alpha, \theta) : \rightarrow R$ such that

$$\left| G(x, w, \alpha, \theta) - f(x) \right| < \varepsilon \quad for\ all\ x \in s \tag{17}$$

$$G(x, w, \alpha, \theta) = \sum_{i=1}^{N} \alpha_j \sigma(w_j^T x + \theta_j) \tag{18}$$

$w_j \in R^n, \alpha_j \in R, \theta_j \in R, w = (w_1, \ldots, w_N),$
$\alpha = (\alpha_1, \ldots, \alpha_N), \theta = (\theta_1, \ldots, \theta_N)$

This theorem can be interpreted as implying that a failure of a function mapping by a multi-layer network, is a result of an inadequate choice of parameters, or of an insufficient number of hidden nodes.

Neural Network Based Controller

A two-layer feedforward N.N of the form

$$r = Z\sigma(Kx) + r_0 \tag{19}$$

is used to approximate the linearizing term u_1 in Eq. (9). Here $r \in R^n$ denotes the network's output, $x = [\ddot{q}m, \dot{q}, q] \in R^{3n}$ is the input that belongs to a compact set, m- number of neurons in the hidden layer $(m > 3n)$, $\sigma(\cdot) \in R^m$ denotes vector of sigmoid functions, $K \in R^{mx3n}$ and $Z \in R^{nxm}$ are the weights of the first and the second layers, respectively, and r_0 is a bias vector.

The function to be approximated by the neural network is given by (see Eq. (9))

$$f = M(q)\ddot{q}_m + V(q, \dot{q}) + G(q) \tag{20}$$

The approximation properties of the neural network is such that

$$\left\| Z\sigma(Kx) + r_0 - (M(q)\ddot{q}_m + V(q, \dot{q}) + G(q)) \right\|_\infty < \varepsilon \tag{21}$$

Combining the neural network control with a PD controller, yields a closed-loop system given by

$$M(q)\ddot{q} + V(q, \dot{q}) + G(q) = r - K_d B(e, \dot{e}) + \tau_d \tag{22}$$

as shown in Fig. 2. Here $B(e, \dot{e}) = \dot{q}(t) - \dot{q}_m(t) + \Lambda(q(t) - q_m(t)) \equiv \dot{e}(t) + \Lambda e$ where $\Lambda = K_d^{-1} K_p \in R^{nxn}$ is a symmetric positive definite matrix, i.e. we assume that K_d and K_p commute. Note, that in most cases one chooses both K_d and K_p to be diagonal matrices and then the motion on the manifold $B(e, \dot{e}) = 0$ is decoupled between the degrees of freedom.

Derivation of Control Law

Define a Lyapunov function

$$E = \frac{1}{2} B^T M B \tag{23}$$

Differentiation of (23) with respect to time yields

$$\dot{E} = B^T (r - v) - B^T (K_d - M\Lambda) B - B^T (M\Lambda - V) \Lambda e \tag{24}$$

where v is the reference to the model system (see Eq. (6)). To ensure stability Lyapunov stability used and the negativity of $\dot{E}$ is ensured under the following conditions. Let $\underline{\alpha}$ and α be the minimum and maximum eigenvalues of Λ, respectively. Then using Eqs. (2)-(5), it can be shown that the domain of attraction is given by

$$D(z) = \begin{cases} \|q(t)\| \le w_1 \\ \|\dot{q}(t)\| \le \frac{\underline{\alpha} \cdot \lambda_{min}(M(q))}{v_b} \end{cases} \tag{25}$$

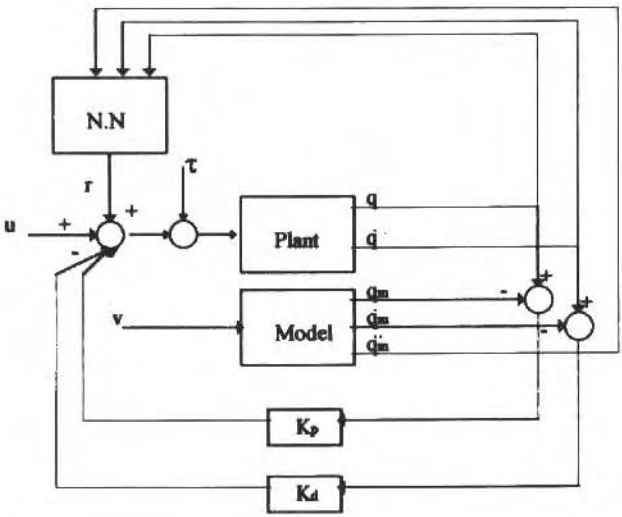


Fig. 2 Closed-loop Configuration of the Proposed Control System

Moreover, let β denote the minimum eigenvalue of K_d, then if

$$\beta > \overline{\alpha} \cdot \lambda_{max}(M(q)) \tag{26}$$

and $\tau_d = 0$, the state converges asymptotically to the mainfold $B = 0$, if

$$B^T(r-v) \le -\| B \| \cdot d \tag{27}$$

where d is determined from initial conditions

$$\| (M - M_m)\ddot{q}_m + (V - V_m) + (G - G_m) \|_{t=t_0} \le d \tag{28}$$

It can be shown that in order to guarantee convergence to the mainfold $B = 0$ in finite time, with $\tau_d \neq 0$, the following conditions need to be satisfied

$$\beta > \overline{\alpha} \cdot \lambda_{max}(M(q)) + \frac{1}{2}\varsigma, \quad \varsigma > 0 \tag{29}$$

$$B^T(r-v) \le -\| B \| (d+\theta) - \eta \cdot (\lambda_{max}(M))^{\frac{1}{2}} (B^T B)^{\frac{1}{2}} \tag{30}$$

Tuning Algorithm

Using a neural network given in Eq. (19) with $r_0 = 0$, the objective is to develop an on-line tuning (learning) algorithm for the weights of the two layers of the neural network, i.e. the elements of matrices K and Z. This is in contrast to many existing neural network-based control approaches that require off-line training.

Tuning of the First Layer

Let $K = K_k$ denote the value of K at time instant k, and denote $K_k^i \in R^{3n}$, $i = 1, ...m$ denote the rows of K_k. Then the matrix K is tuned as follows.

(i) *Initialization step* For $k = 0$, let $K_0 = K^*$, where K^* is an initial value of K.

(ii) *Tuning algorithm* Let Δ_i, $i = 1, ...m$ be a set of predetermined constants, and let $\xi > 1$. Then the tuning is performed according to

$$K_{k+1}^i = \begin{cases} \xi K_k^i & \| K_k^i x \| < \Delta_i \\ K_k^i & otherwise \end{cases} \quad i = 1,...m, k = 1,... \tag{31}$$

Tuning of the Second Layer

Let $Z = Z_k$ denote the value of Z at time instant k. The algorithm for updating Z is derived by solving the following constrained minimization problem is solved

$$\min_{Z} \| r_k - Z_{k-1}\sigma(K_{k-1}x_{k-1}) - K_d B_k \|_2^2 \tag{32}$$

subject to

$$B^T(r_k - v_k) \le -\| B_k \| (d+\theta) - \eta(\lambda_{max}(M))^{\frac{1}{2}} (B_k^T B_k)^{\frac{1}{2}} \tag{33}$$

To solve the problem defined in Eqs. (32) and (33) a discrete version of the gradient projection algorithm is employed [Fletcher (1987)]. The resulting updating equations are as follows:

(i) For $k = 0$, let $Z_0 = Z^*$, where Z^* is an initial value of Z.

(ii) Denote the region in which (33) holds by S. Then the updating of Z_k is performed according to the following rule:

$$Z_{k+1} = \begin{cases} r_k y_{k+1}^+ - \bar{\Gamma}(Z_k y_{k+1} - L_{k+1}) y_{k+1}^+ & \text{if } r_k \in S \\ r_k y_{k+1}^+ - \bar{\Gamma}(Z_k y_{k+1} - L_n) y_{k+1}^+ + \bar{\Gamma} \frac{B_{k+1} B_{k+1}^T}{B_{k+1}^T \bar{\Gamma} B_{k+1}} \bar{\Gamma}(Z_k y_{k+1} - L_{k+1}) y_{k+1}^+ & \text{otherwise} \end{cases}$$

In addition to the advantage of being performed on-line, the algorithm has the following advantages:

a. The weights of Z and K are bounded.

b. As can be seen from the objective function given in Eq. (32), the control effort is minimized.

c. By employing the gradient projection algorithm to solve the constrained minimization problem, the parameter drift problem that occurs in many other algorithms is eliminated. This is due to the fact that the optimization method restricts the network weights to belong to a convex set.

d. Finally, large transient errors common to adaptive systems, are eliminated.

Example: Tracking Control of a Two Degrees of Freedom System

Consider a model of a two degree of freedom system with rotational joints shown in Fig. 3. For this system we have

$$M_{11}(q) = (m_1 + m_2) a_1^2 + m_2 a_2^2 + 2 m_2 a_1 a_2 \cos\theta_2$$
$$M_{12}(q) = M_{21}(q) = m_2 a_2^2 + m_2 a_1 a_2 \cos\theta_2$$
$$M_{22}(q) = m_2 a_2^2$$

$$V(q, \dot{q}) = \begin{bmatrix} -m_2 a_1 a_2 (2\dot{\theta}_1 \theta_2 + \dot{\theta}_2^2) \sin\theta_2 \\ m_2 a_1 a_2 \dot{\theta}_1^2 \sin\theta_2 \end{bmatrix}$$

$$G(q) = \begin{bmatrix} (m_1 + m_2) g a_1 \cos\theta_1 + m_2 g a_2 \cos(\theta_1 + \theta_2) \\ m_2 g a_2 \cos(\theta_1 + \theta_2) \end{bmatrix}$$

$$\tau_d = \left[\sin(t) \;\; \cos(2t) \right]^T$$

It is assumed that the model system is given by

$$M_{m_{11}}(q) = m_{1_m} a_{1_m}^2 \,, \; M_{m_{22}}(q) = m_{2_m} a_{2_m}^2$$

$$M_{m_{12}}(q) = M_{m_{21}}(q) = 0$$

$$V_m(q,\dot{q})=\begin{bmatrix}0\\0\end{bmatrix} \qquad G_m(q_m)=\begin{bmatrix}m_{1_m} g a_{1_m} \cos\theta_1\\ m_{2_m} g a_{2_m} \cos\theta_2\end{bmatrix}$$

The parameters used for the simulation are:

$$m_1=1 \qquad m_2=2 \qquad a_1=2 \qquad a_2=1$$
$$m_{1_m}=1 \qquad m_{2_m}=1.5 \qquad a_{1_m}=3 \qquad a_{2_m}=2$$

The simulation was performed for system with no inputs $(u=\bar{v}=0)$. The neural network used here consisted of two layers with 30 nodes in the hidden layer and a sigmoid was used as an activation function.

Simulation results are shown in Figs. 4-7. As can be observed, both plant outputs converge to the prescribed trajectory while keeping the control effort bounded. It can be also seen that there exists a residual trajectory inaccuracy which is due in the error in the approximation of the linearizing control term, u_1 by the neural network. This error can be reduced by increasing the number of nodes at each layer.

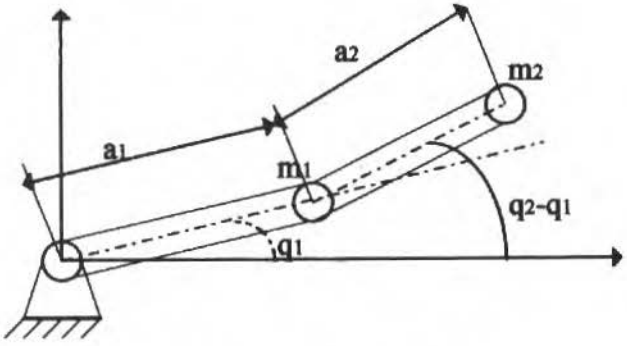


Fig. 3 A Two Degrees of Freedom System

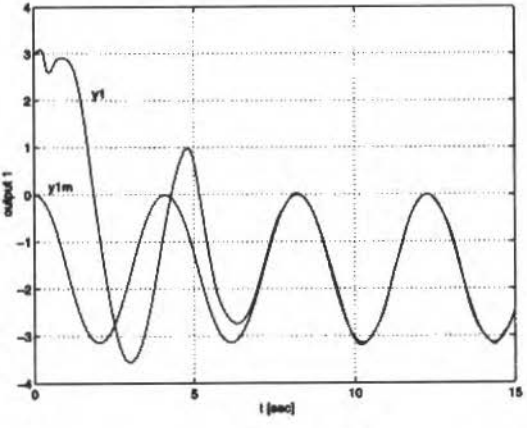


Fig. 4 Simulation Study: Transient Behavior of $\theta_1(t)$

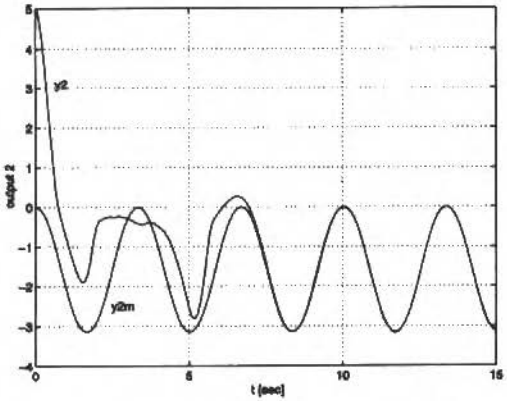


Fig. 5 Simulation Study: Transient Behavior of $\theta_2(t)$

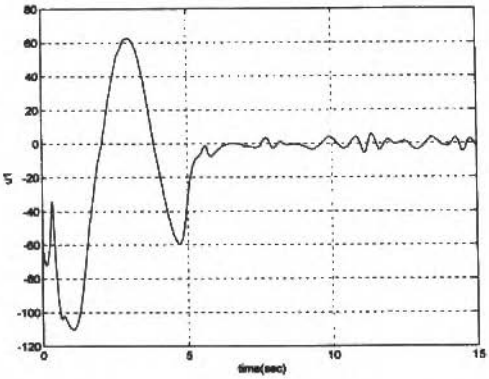


Fig. 6 Simulation Study: Control Effort $u_1(t)$

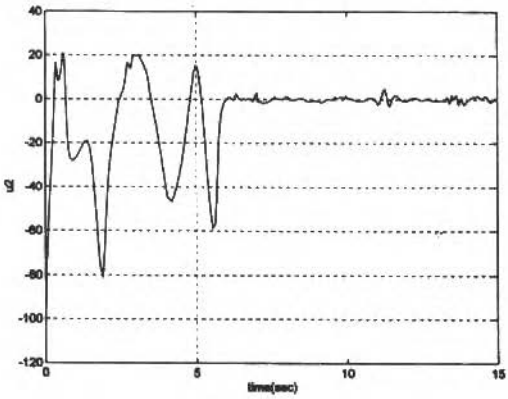


Fig. 7 Simulation Study: Control Effort $u_2(t)$

Conclusions

A neural network-based method for tracking control of mechanical system was developed. The controller consists of a proportional and derivative control action, and a neural network employed to approximate the linearizing term of the controller. Tuning the neural network weights is performed by solving a constrained convex optimization problem whose solution is well understood. It is shown that the control law causes the given system to follow a model trajectory without assuming knowledge of the system's parameters, and in presence of disturbances. The resulting algorithm is very simple and requires a modest computing effort. The algorithm has the advantage of adjusting the weights on-line. The proposed approach avoids problems common to adaptive systems, such as parameter drift and large errors during transient behavior. Finally, these tasks are achieve while minimizing the control effort.

References

Chernousko, F.L., 1996, "Control of Nonlinear Mechanical Systems", Proceedings of the 2nd European Nonlinear Oscillations Conference, Prague, September 19-13.

Chernousko, F.L. 1, 1993, "The Decomposition of Controlled Dynamic Systems", in Advances in Nonlinear Dynamics and Control: A Report from Russia, A.B. Kurzhanski, Editor, Birhauser, Boston, pp. 1-40.

Corless, M., 1993, "Control of Uncertain Nonlinear Systems", Transactions of the ASME, Vol. 115, pp. 362-372.

Cybenko, G., 1989, "Approximation by Super-Position of a Sigmoidal Functions", Math. of Control, Signals and Systems, Vol. 2, pp. 303-314.

Decarlo, R.A., Zak, S.H. and Matthews, G.P., 1988, "Variable Structure Control of Nonlinear Multivariable Systems", IEEE Proc., Vol. 76, No. 3, pp. 212-232.

Fletcher, R., 1987, Practical Methods of Optimization, second edition, John Wiley & Sons.

Hassoum, M.H., 1995, Fundamentals of Artificial Neural Networks, MIT Press.

Jung, S. and Hsia, T.C., 1995, "A New Neural Network Control Technique for Robot Manipulators", Robotica, Vol. 13, pp. 477-484.

Khemaissia, S. and Morris, A.S., 1993, "Neuro-Adaptive Control for Robotic Manipulators", Robotica, Vol. 11, pp. 465-473.

Lewis, F.L., Abdallah, C.T. and Dawson, D.M., 1993, "Control of Robot Manipulators", Macmillan.

Lewis, F.L., Liu, K. and Yesildirek, A., 1995, "Neural Nets Robot Controller with Guaranteed Tracking Performance", IEEE Transactions on Neural Networks, Vol. 6, No. 3, pp. 703-715.

Lewis, F.L., Liu, K. and Yesildirek, A., 1996, "Multilayer Neural Net Robot Controller with Guarantee Tracking Performance", IEEE Transactions on Neural Networks, Vol. 7, No. 2, pp. 388-399.

Morris, A.S. and Khemaissia, 1995, "Stable and Fast Neurocontroller for Robot Arm Movement", IEEE Proc. Control Theory Appl., Vol. 142, No. 4, pp. 378-384.

Narendra, K.S. and Parthasarathy, K., 1991, "Gradient Methods for the Optimization of Dynamical Systems Containing Neural Networks", IEEE Transactions on Neural Networks, Vol. 2, pp. 252-262.

Narendra, K.S. and Parthasarathy, K., 1990, "Identification and Control of Dynamical Systems Using Neural Networks", IEEE Transactions on Neural Networks, Vol. 1, pp. 4-27.

Sarangapani, J. and Lewis, F.L., 1990, "Discrete-Time Model Reference Adaptive Control of Nonlinear Dynamical Systems Using Neural Networks", Int. J. Control, Vol. 64, No. 2, pp. 217-239.

Slotine, J-J.E. and Li, W., 1991, Applied Nonlinear Control, Prentice Hall.

Slotine, J-J.E. and Li, W., 1988, "Adaptive Manipulator Control: A Case Study", IEEE Trans. on Automat. Control, Vol. 33, No. 11, pp. 995-1003.

Spong, E. and Vidyasgar, M., 1989, Robot Dynamics and Control, John Wiley.

Sanner, R.M. and Slotine, J-J.E., 1995, "Stable Adaptive Control of Robot Manipulators Using Neural Networks", Neural Computation, Vol. 7, pp. 753-790.

Utkin, V.I., Sliding Modes and Their Application in Variable Structure Systems, Moscow, MIR Publishers.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 228-241

ISSN 0100-7386
Printed in Brazil

A Consistent Approach to Treat the Dynamics of Flexible Systems

Fernando Alves Rochinha
Universidade Federal do Rio de Janeiro - EE/COPPE
Departamento de Engenharia Mecânica
21945-970 Rio de Janeiro, RJ Brasil

Rubens Sampaio
Pontifícia Universidade Católica do Rio de Janeiro
Departamento de Engenharia Mecânica
22453-900 Rio de Janeiro, RJ Brasil

Abstract

The dynamics of flexible systems, such as robot manipulators, mechanical chains or cables, is becoming increasingly important in engineering. The main question arising from the numerical modeling of large displacements of multibody systems is an appropriate treatment for the large rotations. In the present work an alternative approach is proposed leading to a time-stepping numerical algorithm which achieve stable solutions combined with high precision. In particular, in order to check the performance of the proposed approach, two examples having preserved constants of the motion are presented.
Keywords: *Flexible Mechanical Systems, Multibody System Dynamics.*

Introduction

Recently there has been a great interest in the study of nonlinear dynamics of structures and its applications to a wide variety of engineering problems: robotics, spacecraft dynamics and attitude control, vehicle dynamics, large structures vibrations. Indeed, there is a growing literature (see, for example, Borri et al., 1991; Hughes, 1986; Vu-Quoc and Simo, 1988; and references therein).

The present article is concerned with the mechanical behavior of multibody flexible systems. The name multibody stands as a general term that encompasses a wide range of systems such as: mechanisms, automobiles and trucks, robots, trains, space structures, antennas, satellites, the human body, and others. In a very broad sense, those systems are modeled as simply collection of bodies with a given connection configuration. The bodies of the system may be either rigid or flexible (often modeled as rods, plates or shells); and they may form closed or open loops. They are linked by means of connections like spherical or pin joints.

The importance of structural flexibility in multibody systems, as in the case of space manipulators, has been recognized by many researches. As a result, the subjects of modeling elasticity and formulating the governing equations for dynamic simulation of these systems have been investigated extensively over the past years. Until recently, however, in order to encompass the flexibility in the modeling of multibody systems the assumption of small strains was adopted. This hypothesis leads to unrealistic approaches (Sharf, 1995, 1996) with serious drawbacks, like, for instance, the lack of geometric stiffening within rotational problems. That situation motivated several researches to develop new methodologies for dynamic analysis of flexible-body systems by accounting for large displacements and finite rotations. For example, Vu-Quoc and Simo (1988), Geradian and Cardona (1989) and Downer et al. (1992) proposed a formulation based on the fully non-linear or geometrically exact theory for rods. Those theories are known as consistent as they do not have any kind of kinematical simplification within the context of one dimensional continuum.

In particular, the approach to the rotational degrees of freedom plays a crucial role in the modeling and numerical approximation using finite element method of rods and shells, which are often used as links in multibody systems (see Vu-Quoc and Simo, 1988; Rochinha, 1990; and Le Tallec et al., 1992). Thus, in order to better establish and assess the performance of some specific approach, the dynamics of a single rigid body is often used as prototype problem. From a numerical standpoint it represents a significant test due to its high degree of nonlinearity.

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors. Agenor de Toledo Fleury and Hans Ingo Weber.

The dynamic response associated with the rotational degrees of freedom leads to an evolution problem in the rotation group, which can be parametrized in several different ways (Hughes, 1986). In the present work, a variational formulation, closely related to that proposed by the Le Tallec et al. (1992), based on the use of directors (Antman and Kenney, 1981) is presented. The use of lagrange multipliers avoids some difficulties in obtaining the tangent operators inherents to numerical procedures and leads to a symmetrical operator as well.

The objective of the present work is to develop a class of numerical algorithms dedicated to the simulation of the rotational nonlinear dynamics of multibody systems involving elastic or rigid links. The elastic links are often modelled by means of a geometric exact one-dimensional rod's theory discussed in Le Tallec et al. (1992). In this theory, the kinematics of the cross section is described using the same framework adopted for the rotational motion of a rigid body. So, the nonlinear dynamics of rigid bodies is a key point in the modeling and simulation of multibody systems. A special attention is devoted to the conservation of fundamental quantities associated to the motion, such, for instance, the total energy or the angular momentum. From the numerical standpoint, the conservation of energy by the integration algorithm is a manifestation of unconditional stability (Simo and Wong, 1994).

It is also worthwhile to stress the importance of an accurated description of the nonlinear dynamics in the control synthesis of multibody systems. Very often, the resulting problems are not amenable to methods of linear control theory, and they are not transformable into linear control problems in any meaningful way. Hence, these are nonlinear control problems that require fundamentally nonlinear approaches. In particular, in the presence of conserved quantities, due to symmetry properties of the motion, a nonholonomic nonlinear control problem may arise (Kolmanovsky and McClamroch, 1995). Thus, in order to better understand the behavior of a controled multibody system, a reliable numerical integrator, which inherits the same features of the original problem plays a crucial role.

Mechanical Modeling

In the present section two nonlinear models are presented. The first deals with rigid links and the other with flexible links of a multibody system. The main idea is to emphasize that both models share some common features in regard to the rotational degrees of freedom.

Let $\P$ be the reference placement of a solid body, with particles labeled by $\mathbf{X}$. So, the motion of the body is described by the mapping:

$$\phi : \P \times [0,T] \rightarrow R^3$$

$$(X,t) \rightarrow \boldsymbol{x} \tag{1}$$

where t denotes the time, T represents the total time of observation and $\mathcal{R}^3$ is the ordinary 3D-Euclidean space. Throughout this paper, bold letters are used to designate vectors.

In the case of rigid bodies, the position of each particle can be rewritten in the following form:

$$\mathbf{x}(t)=\mathbf{r}(t)+\sum_{i=1}^{3} X_i \mathbf{d}_i(t) \tag{2}$$

where $\mathbf{r}$ defines the position of the center of mass and $\mathbf{d_i}$ are vector fields called directors (see Antman and Kenney, 1981), forming at each instant t an orthonormal basis attached to the body. Within the rigid body literature, the basis $\{\mathbf{d_i}\}$ is known as the body frame.

In (2), **r** takes into account the translation of the body and the comparison among the directors and a fixed basis gives the rotation of the body. Thus, the configuration of the body is defined by 6 parameters forming the pair

$$(r, \mathbf{d_i}) \in R^3 \times K = \{ R^9 / \mathbf{d_i}.\mathbf{d_j} = \delta_{ij} \} \tag{3}$$

where δ_{ij} represents the delta of kronecker. Consequently, one refers to R^3 x K as the abstract configuration manifold of the rigid body.

Remark: Indeed, the manifold *K* and the special group of rotations *SO*(3) can be identified by the isomorphism:

$$\mathbf{d_i} = \Lambda \mathbf{D_i} \quad (i = 1,3) \tag{4}$$

where Λ is a rotation belonging to *SO*(3) and $\mathbf{D_i}$ are the directors in the reference placement, which, without lost a generality, can be taken coincident with the principal directions of inertia.

The velocities and accelerations fields in the present theory are

$$\mathbf{v} = \dot{\mathbf{r}} \quad ; \quad \mathbf{a} = \dot{\mathbf{v}} \quad ; \quad \dot{\mathbf{d}}_i \quad ; \quad \ddot{\mathbf{d}}i$$

where $\dot{f}$ stands for the time derivative of *f*. The first two fields are, respectively, the translational velocity and the translational acceleration. The others describe the velocity and acceleration of the angular motion and are related to angular velocity **w** and angular acceleration α of the body by means of:

$$\dot{\mathbf{d}}_i = \mathbf{w} \Lambda \mathbf{d}_i$$

$$\ddot{\mathbf{d}}_i = \alpha \Lambda \mathbf{d}_i + \mathbf{w} \Lambda \dot{\mathbf{d}}_i$$

where Λ denotes the ordinary vector product.

Within the same context, a geometric exact nonlinear formulation for rods can be developed. The both presented models use the same approach for the description of finite rotations (rotational degrees of freedom). Before introducing the formal modelling a conceptual overview of a rod is discussed. The rods, in which the independent variables are the arclenght *S* and the time t, is represented by a curve in space, with a right-handed orthonormal frame of directors attached at each point along the curve. The curve describes the location of the centerline of the rod. The triad of directors describes the orientation of the material cross section of the rod.

The kinematics of the rod is summarized in the following expression:

$$\mathbf{x}(t) = \mathbf{r}(S,t) + \sum_{i=1}^{2} X_i \, \mathbf{d_i}(S,t) \tag{5}$$

where **r** defines he position of the center of the cross section and $\mathbf{d_i}$ *(i = 1, 2)* are orthonormal vector fields called directors defining, at each pair *(S,t)*, the plane of the cross section. Introducing a third director, $\mathbf{d_3}$, perpendicular to the other two and, consequently, perpendicular to the cross section, the deformation of the rod, flexure and torsion, is described by the relative orientation of the triad (Antman and Kenny, 1981). From this perspective, the set of admissible configurations, for a fixed *t*, of a rod is defined by:

$$Q_t = \{ \phi := (\mathbf{r}, \mathbf{d_i}) : [0, L] \rightarrow R^3 \times K \text{ and } \phi \text{ satisfying appropriated boundary conditions} \} \tag{6}$$

where 0 and L defines the extremities of the rod.

Comparing the definitions contained in (3) and (6) it turns out the correspondence between the two models.

Variational Principle: Equations of Motion

In the present work, the equations of motion of a rigid body will be obtained by means of the postulation of a variational principle, which is nothing but the Least Action Principle (see Meirovitch, 1970; and Borri et al., 1991) in the present context. The extension of the proposed mathematical formulation for nonlinear rods is presented in Campos (1996).

Based on the general formula of the kinetic energy of a continuum, an expression for the kinetic energy in the present modelling of rigid body is obtained using (2)

$$T=\int_S \frac{1}{2}\rho\dot{\mathbf{x}}.\dot{\mathbf{x}}\,dS=\frac{1}{2}\{\int_S \rho(\mathbf{v}+\sum_{i=1}^{3}X_i\dot{\mathbf{d}}_i)^2\,dS\}=$$

$$\frac{1}{2}\{\int_S \rho\mathbf{v}.\mathbf{v}dS+\int_S \rho\sum_{i=1}^{3}X_i^2\,\dot{\mathbf{d}}_i.\dot{\mathbf{d}}_i dS+\int_S \rho\sum_{i=1}^{3}\sum_{j=1}^{3}X_iX_j\dot{\mathbf{d}}_i.\dot{\mathbf{d}}_j dS\}$$

where ρ denotes the mass density of the body.

Using the well known definitions of inertia moments, the expression above is rephrased yielding

$$T=\frac{1}{2}M\mathbf{v}.\mathbf{v}+\frac{1}{2}\{\{\frac{I_{33}-I_{11}+I_{22}}{2}\}\dot{\mathbf{d}}_1.\dot{\mathbf{d}}_1+\{\frac{I_{11}-I_{22}+I_{33}}{2}\}\dot{\mathbf{d}}_2.\dot{\mathbf{d}}_2+$$

$$\{\frac{I_{22}-I_{33}+I_{11}}{2}\}\dot{\mathbf{d}}_3.\dot{\mathbf{d}}_3-2I_{12}\dot{\mathbf{d}}_1.\dot{\mathbf{d}}_2-2I_{13}\dot{\mathbf{d}}_1.\dot{\mathbf{d}}_3-2I_{23}\dot{\mathbf{d}}_2.\dot{\mathbf{d}}_3\}$$

where M is the total mass of the body and I_{ij} are the inertia moments.

The potential energy associated to the external loads is given by

$$V=-\mathbf{f}.\mathbf{r}-\mathbf{f}_1.\mathbf{d}_1-\mathbf{f}_2.\mathbf{d}_2-\mathbf{f}_3.\mathbf{d}_3 \tag{7}$$

where the $\mathbf{f}_i$ ($i = 1, 3$) are implicitly defined by

$$\mathbf{m}=\mathbf{d}_1\wedge\mathbf{f}_1+\mathbf{d}_2\wedge\mathbf{f}_2+\mathbf{d}_3\wedge\mathbf{f}_3 \tag{8}$$

and where $\mathbf{m}$ and $\mathbf{f}$ are, respectively, the total torque and the resultant force applied in the center of mass.

So, from the Lagrangian $L = T - V$ defined over $R^3 \times K$, the following variational principle associated with the Least Action Principle is obtained

$$\int_1^2\{m\mathbf{a}.\mathbf{p}-\mathbf{f}.\mathbf{p}\}\,dt+\int_1^2\{\{\frac{I_{33}-I_{11}+I_{22}}{2}\}\ddot{\mathbf{d}}_1.\mathbf{g}_1+$$

$$\{\frac{I_{33}-I_{11}+I_{22}}{2}\}\ddot{\mathbf{d}}_2.\mathbf{g}_2+\{\frac{I_{22}-I_{33}+I_{11}}{2}\}\ddot{\mathbf{d}}_3.\mathbf{g}_3-I_{12}\ddot{\mathbf{d}}_2.\mathbf{g}_1$$

$$- I_{12}\ddot{\mathbf{d}}_1 \cdot \mathbf{g}_2 - I_{13}\ddot{\mathbf{d}}_3 \cdot \mathbf{g}_1 - I_{13}\ddot{\mathbf{d}}_1 \cdot \mathbf{g}_3 - I_{23}\mathbf{g}_2 \cdot \ddot{\mathbf{d}}_3 \} - I_{23}\ddot{\mathbf{d}}_2 \cdot \mathbf{g}_3 \}$$

$$-\mathbf{f}_1 \cdot \mathbf{g}_1 - \mathbf{f}_2 \cdot \mathbf{g}_2 - \mathbf{f}_3 \cdot \mathbf{g}_3 = 0$$

$$\forall (\mathbf{p}, \mathbf{g}_i) \in R^3 \, x \, dK \tag{9}$$

where dK is the tangent space to K given by

$$dK = \{(p, g_i) \in R^9 \, x [t \ , t \ ] \, ; \, g_i = U \Lambda d_i, g_i(t_1) = g_i(t_2) = 0 \text{ and } p(t_1) = p(t_2) = 0\}$$

where $[t_1, t_2] \subset [0,T]$

Remark: The variational Eq. (9) can be shown to be equivalent to the classical equations of rigid body dynamics by using standard arguments of calculus of variation and choosing: $\mathbf{g_i} = \mathbf{U} \Lambda \mathbf{d_i}$.

For numeral purposes, the form of (9) is rephrased by using lagrange multipliers, yielding

$$\int_{t1}^{t2} \{m\mathbf{a}.\mathbf{p} - \mathbf{f}.\mathbf{p}\} dt + \int_{t1}^{t2} \{\{\frac{I_{33} - I_{11} + I_{22}}{2}\} \ddot{\mathbf{d}}_1 \cdot \mathbf{g}_1 +$$

$$\{\frac{I_{11} - I_2 + I_{33}}{2}\} \ddot{\mathbf{d}}_2 \cdot \mathbf{g}_2 + \{\frac{I_{22} - I_{33} + I_{11}}{2}\} \ddot{\mathbf{d}}_3 \cdot \mathbf{g}_3$$

$$- 2I_{12}\ddot{\mathbf{d}}_2 \cdot \mathbf{g}_1 - 2I_{12}\ddot{\mathbf{d}}_1 \cdot \mathbf{g}_2 - 2I_{13}\ddot{\mathbf{d}}_3 \cdot \mathbf{g}_1 - 2I_{13}\ddot{\mathbf{d}}_1 \cdot \mathbf{g}_3 - 2I_{23}\mathbf{g}_2 \cdot \ddot{\mathbf{d}}_3 - 2I_{23}\ddot{\mathbf{d}}_2 \cdot \mathbf{g}_3 +$$

$$-\sum_{i=1}^{3}\sum_{j=1}^{3} \lambda_{ij} \left\{\mathbf{d}_i \cdot \mathbf{g}_j + \mathbf{d}_j \cdot \mathbf{g}_i\right\} - \sum_{i=1}^{3} f_i \cdot g_i dt = 0$$

$$\forall (\mathbf{p}, \mathbf{g}_i) \in R^3 \, x \, R^9 \tag{10}$$

where λ_{ij} are the lagrange multipliers associated with the orthonormality of the directors and, consequently, to the rigid body condition.

Numerical Algorithm

The translational part of the variational problem (10) can be approximated by means of any standard integrator as, for example, the Newmark algorithm, which details will not be included here. The rotational part is equivalent to a non-linear evolution problem on S0(3), which entails some difficulties (see Geradin and Cardona, 1989). So, in the present work it is proposed the following evolution algorithm in connection with the Newton method:

Let $[t_n, t_{n+1}] \subset [0,T]$ be a typical interval where: $[0,T] = U_{n=0}^{N} [t_n, t_{n+1}]$.

Let $\Delta t = t_{n+1} - t_n$ be the time step.

Assume that at t_n the following initial data are known:

$$(\mathbf{d}_{i_n}, \dot{\mathbf{d}}_{i_n}) \in K + R^9$$

The objective is to obtain an approximation

$$(\mathbf{d}_{i_{n+1}}, \dot{\mathbf{d}}_{i_{n+1}}, \lambda_{ij_{n+\frac{1}{2}}}) \in K + R^9 \times R^9$$

to the actual solution $(\mathbf{d}_i(t_n), \dot{\mathbf{d}}_i(t_{n+1}), \lambda_{ij}(t_n + \frac{\Delta t}{2}))$. The main steps of the algorithm are summarised below.

Step 0. Initialization for time step in $[t_n, t_{n+1}]$.

- Define a predictor for directors and velocities:

$$\mathbf{d}^0_{i_{n+1}} = \mathbf{d}_{i_n} \quad ; \quad \dot{\mathbf{d}}^0_{i_{n+1}} = -\dot{\mathbf{d}}_{i_n}$$

Step1. Compute the Lagrange multipliers $\lambda_{ij^t_{n+\frac{1}{2}}}$

(i=0,...,N-1) by solving

$$I_A \{\dot{\mathbf{d}}^t_{1_{n+1}} \cdot \mathbf{g}_{1_{n+1}} - \dot{\mathbf{d}}_{1_n} \cdot \mathbf{g}_{1_n}\}$$

$$+ I_B \{\dot{\mathbf{d}}^t_{2_{n+1}} \cdot \mathbf{g}_{2_{n+1}} - \dot{\mathbf{d}}_{2_n} \cdot \mathbf{g}_{2_n}\}$$

$$I_C \{\dot{\mathbf{d}}^t_{3_{n+1}} \cdot \mathbf{g}_{3_{n+1}} - \dot{\mathbf{d}}_{3_n} \cdot \mathbf{g}_{3_n}\}$$

$$-\frac{1}{2}\Delta t \{\sum_{i,j=1}^{3} \lambda_{ij_{n+\frac{1}{2}}} \{\{\mathbf{d}^t_{i_{n+1}} \cdot \mathbf{g}_{j_{n+1}} + \mathbf{d}^t_{j_{n+1}} \cdot \mathbf{g}_{i_{n+1}}\}$$

$$+ \{\mathbf{d}_{i_n} \cdot \mathbf{g}_{j_n} + \delta_{ij} \cdot \mathbf{g}_{i_n}\}\}\}$$

$$-\Delta t \sum_{i=1}^{3} \frac{\mathbf{f}_{i_{n+1}} + \mathbf{f}_{i_n}}{2} \cdot \mathbf{g}_i = 0$$

for a convenient choice of $\mathbf{g_i}$. Where $I_A = \frac{I_{33} - I_{11} + I_{22}}{2}$, $I_B = \frac{I_{11} - I_{22} + I_{33}}{2}$, $I_C = \frac{I_{22} - I_{33} + I_{11}}{2}$. In the above expression, the directors are taken coincident with the principal directions of inertia.

Step 2. Compute residual R^i_{n+1} of the above equation and check convergence

- Computation of R^i_{n+1} by choosing $g_i = U \Lambda d^i_{i_{n+1}}$.

- Check for convergence: $IF \left\| R^i_{n+1} \right\| \leq$ tolerance then begin a new time step ($n+1 \rightarrow n$; go to Step 0). ELSE continue

Step 3. Compute the tangent matrix K^i_{n+1} and solve for the rotation increment $\hat{g}_i = \hat{U} \Lambda d_i \in K$

- Computation of K^i_{n+1}

$$K^i_{n+1} = \frac{2}{\Delta t} \begin{pmatrix} I_A & 0 & 0 \\ 0 & I_B & 0 \\ 0 & 0 & I_C \end{pmatrix}$$

$$- \frac{\Delta t}{2} \begin{pmatrix} 2(\lambda_{22^i}{}_{n+\frac{1}{2}} + \lambda_{33^i}{}_{n+\frac{1}{2}}) & -\lambda_{12^i}{}_{n+\frac{1}{2}} & -\lambda_{13^i}{}_{n+\frac{1}{2}} \\ -\lambda_{12^i}{}_{n+\frac{1}{2}} & 2(\lambda_{11^i}{}_{n+\frac{1}{2}} + \lambda_{33^i}{}_{n+\frac{1}{2}}) & -\lambda_{23^i}{}_{n+\frac{1}{2}} \\ -\lambda_{13^i}{}_{n+\frac{1}{2}} & -\lambda_{23^i}{}_{n+\frac{1}{2}} & 2(\lambda_{11^i}{}_{n+\frac{1}{2}} + \lambda_{22^i}{}_{n+\frac{1}{2}}) \end{pmatrix}$$

- Computation of $\hat{\mathbf{U}}$ solving the linear system

$$K^i_{n+1} \hat{\mathbf{U}} = R^i_{n+1}$$

Step 4. Update the configuration and the velocity for a given increment $\hat{\mathbf{g}}_\mathbf{i}$

- $\mathbf{d}^{i+1}_\mathbf{i} = Proj_{SO(3)}(\mathbf{d}^i_\mathbf{i} + \hat{\mathbf{g}}_\mathbf{i}) = \Lambda^{i+1}_n \mathbf{d}_{\mathbf{i}_n}$

- $\mathbf{d}^{i+1}_\mathbf{i} = \Lambda^{i+1}_n (\frac{2}{\Delta t} \theta \wedge \mathbf{d}_{\mathbf{i}_n} - \mathbf{d}_{\mathbf{i}_n})$

where $Proj_{SO(3)}$ is a projection into K, adopted here as exponential transform (see Simo and Wong, 1994 or Hughes, 1986) and $\Lambda^{i+1}_n = exp\,\theta$, with exp denoting the exponential operator. Indeed Λ^{i+1}_n plays the role of a transport between the tangent spaces to the configurations assumed at t_n and t_{n+1} (iteration $i+1$). In the above expressions the subscripts $n+1$ were omitted.

Step 5. Begin a new iteration; $i+1 \rightarrow i$; go to Step2.

Numerical Simulations

In this section, two representative numerical simulations are presented in order to illustrate the performance of the proposed algorithm. The both situations, which are also examined by Simo and Wong (1991) and Park and Chiou (1993), are choosen due to the presence of conserved quantities

during the motion. In fact, the reproduction of this conservation behavior by the numerical algorithm is considered a demonstration of stability and accuracy. The key idea is to treat problems in which the reliability of the proposed numerical modelling can be assessed even in the long time dynamics.

In the first simulation, it is considered the motion of a symmetrical top with total mass M in a uniform gravitational field $-g\,\mathbf{e}_3$. The vector $\mathbf{e}_3$ belongs to an inertial frame $\{\mathbf{e}_i\}$. The same numerical values of Simo and Wong (1991) were adopted: $M_g = 20$, $l = 1$, $I_{11} = I_{22} = 5$, and $I_{33} = 1$. The parameter l denotes the distance between the bottom of the top and its center of gravity. The following initial conditions are also adopted:

$$\mathbf{w}(0)=[0\ 0\ 50]; \qquad \mathbf{d}_i(0)=\exp([0.3\ 0\ 0])\mathbf{e}_i$$

Figures 1-4 shows the numerical results obtained for time step $\Delta t = 0.001$. The total energy remains constant during the observed period as well as the component of the spatial angular momentum in the direction $\mathbf{e}_3$, h_3. Both facts demonstrate the stability and accuracy of the proposed formulation in the present example as this conservation behavior is to be expected. Figure 4 presents the nutation and precession of the top about its fixed contact point. For the present example (see Simo and Wong, 1991), one has the following relations for the angular frequencies of nutation w_n and precession w_p:

$$w_n = \frac{I_{33}}{I_{11}}(\mathbf{w}.\mathbf{d}_3) \qquad \text{and} \qquad w_p = \frac{Mgl}{I_{33}(\mathbf{w}.\mathbf{d}_3)}$$

For the choosen numerical parameters these relations lead to value $w_n = 10.0$ and $w_p = 0.4$. In the present numerical simulation $w_n = 9.9999$ and $w_p = 0.4000$ were obtained.

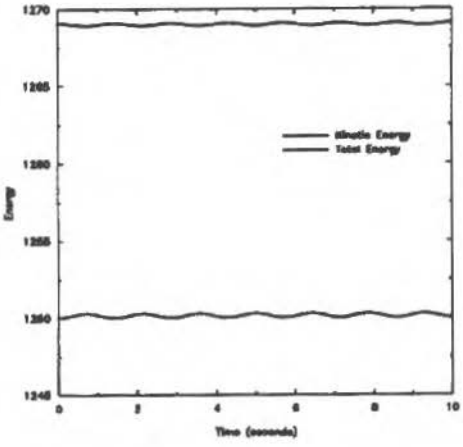


Fig.1 Kinetic and Total Energy

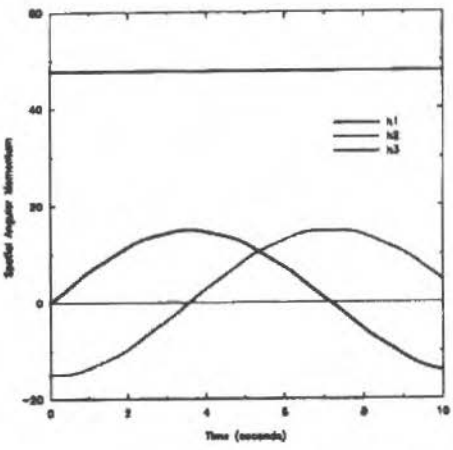


Fig. 2 Spatial Angular Momentum

The numerical solution obtained with $\Delta t = 0.001$ will be considered a good approximation to the true solution, which is not known, in order to assess the main features of the proposed numerical scheme. In figures 5, 6 and 7 the numerical solution using $\Delta t = 0.04$ (40 times the first one) is depicted. A constant oscillation is exhibited by the total energy, although it does not represent a signal of numerical instability as it is minute and constant in the long period of observation. It is also important to observe the preservation of the component h_3 of the Spatial Angular Momentum. The values obtained for the angular velocities of nutation and precesion are, respectively, $w_n = 9.5460$ and $w_p = 0.4190$.

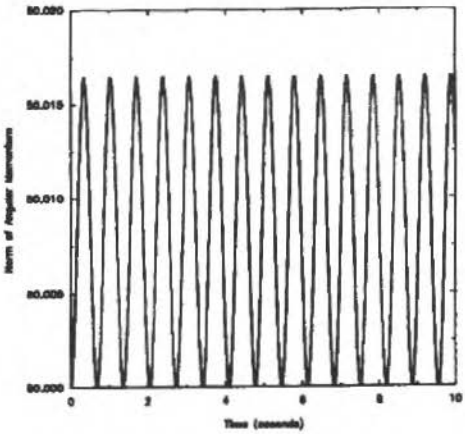


Fig. 3 Norm of Angular Momentum

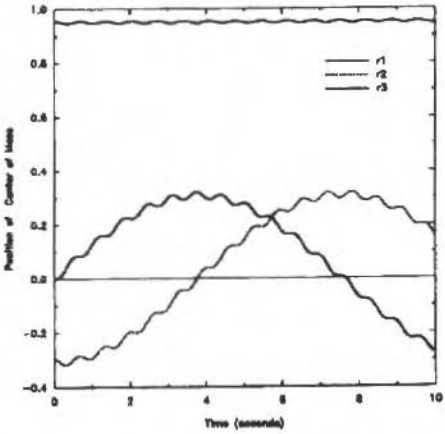


Fig. 4 Position of Center of Mass

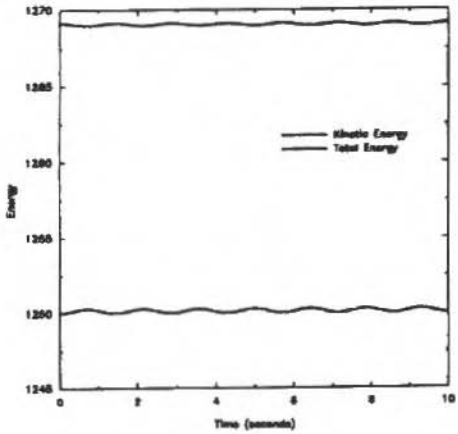


Fig. 5 Kinetic and Total Energy ($\Delta t = 0.04$)

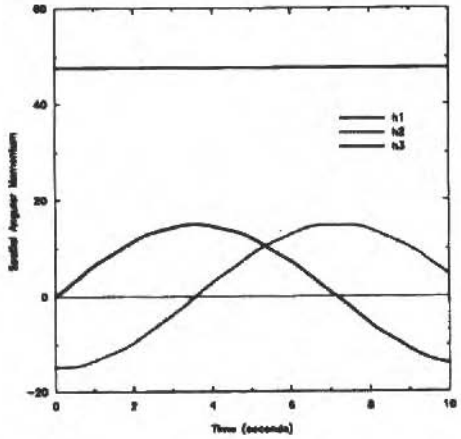


Fig. 6 Spatial Angular Moment *(Δt = 0.04)*

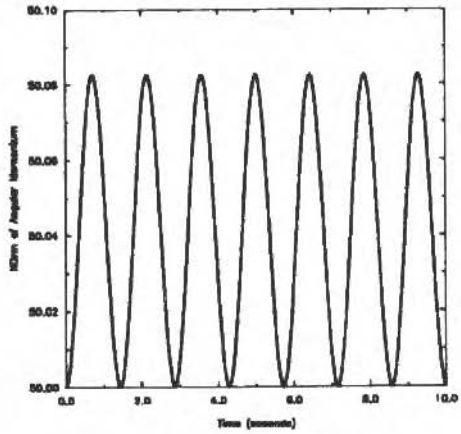


Fig. 7 Norm of Angular Momentum *(Δt = 0.04)*

The second numerical experiment deals with the unstable rotational motion about the intermediate moment of inertia of a rigid body. This simulation is performed according to the three following load steps: (i) At time $t = 0$ a constant torque is applied to the body at rest in the direction coincident with the intermediate moment of inertia; (ii) At time $t = t^*$, the torque is removed and another one is applied to the body for a short duration equal to the time step Δt; (iii) Finally, the rigid body undergoes a torque free motion. The situation is summarized in the torque history given below:

$$\mathbf{m} = \begin{cases} A_1\mathbf{e}_1 & 0 \le t \le t^* \\ A_2\mathbf{e}_2 & t^* \le t \le t^* + \Delta t \\ 0 & t > t^* + \Delta t \end{cases}$$

where A_1 and A_2 are constants given by: $A_1 = 20$ and $A_2 = \frac{0.2}{\Delta t}$. The another parameters used in the simulations are: $t^* = 2.0 - \Delta t$, $I_{11} = 5$, $I_{22} = 10$ and $I_{33} = 1$. Although those inertia moments do not correspond to a real rigid body, as the sum of two of them, namely: I_{11} and I_{33}, is not greater than

the third one, they were taken to make comparisons with the results presented in Simo and Wong (1991) and Park and Chiou (1993).

The body begins at rest, thus initial velocities are zero. As the result force acting on the body is zero during the motion, the center of mass stays at rest. Hence, this is a pure rotational situation in which the angular momentum and the total energy are conserved after the instant $t^* + \Delta t$.

Two distincts time steps were choosen for the numerical experiments, namely: 0.1 and 0.01. The solution obtained using the second time step will be taken as a reference and it is summarized in Figs. 8-11. The first three, kinetic energy, the norm of angular momentum and the spatial angular momentum, demonstrate the good conservation capabilities of the proposed algorithm. In the last one, Fig. 11, the convected angular velocity (see Simo and Wong, 1991) is depicted.

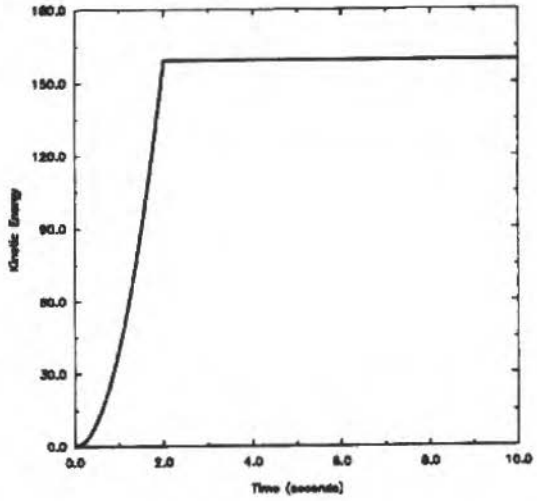


Fig. 8 Kinetic Energy

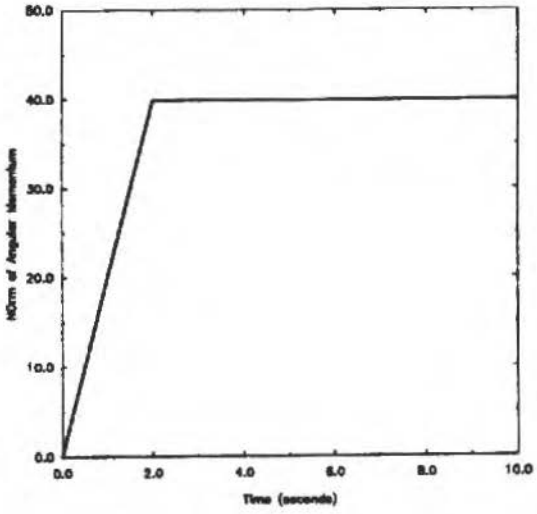


Fig. 9 Norm of Angular Momentum

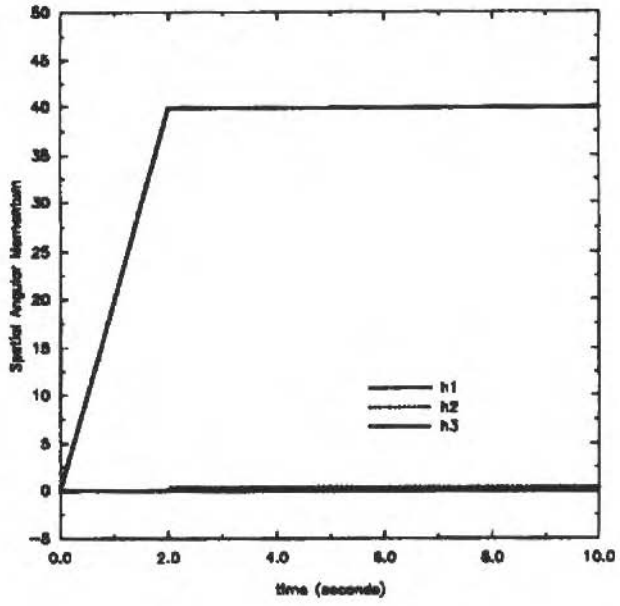


Fig.10 Spatial Angular Momentum

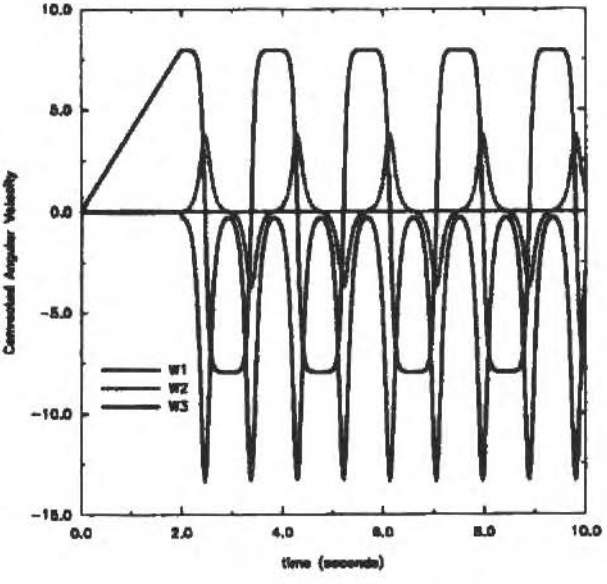


Fig.11 Convected Angular Velocity

The results obtained using the $\Delta t = 0.1$ are presented in figures 12-15. In the present modelling, although the applied moment introduced above remains constant during the intervals $[0,t^*]$ and $[t^*,t^*+\Delta t]$, for this case the applied load depends on the assumed configuration due to the definition 8. This fact explains the discrepancy observed between the results using the two different time steps.

Indeed, $\Delta t = 0.1$ implies in rather large values of incremental rotations and leads to inaccurate results. Nevertheless, the energy and angular momentum are still conserved.

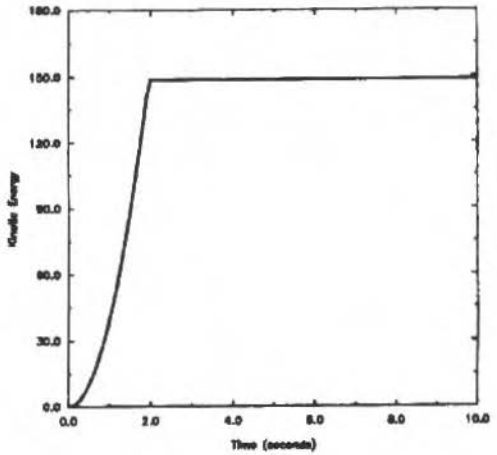


Fig. 12 Kinetic Energy

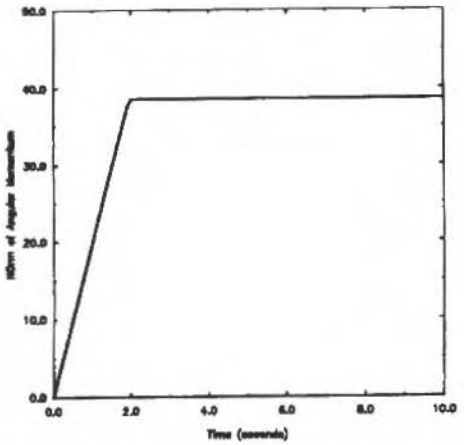


Fig. 13 Norm of Angular Momentum

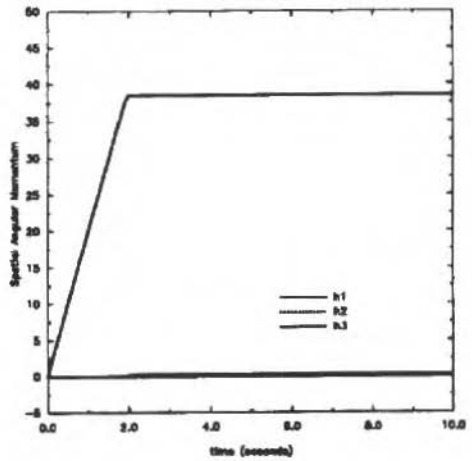


Fig. 14 Spatial Angular Momentum

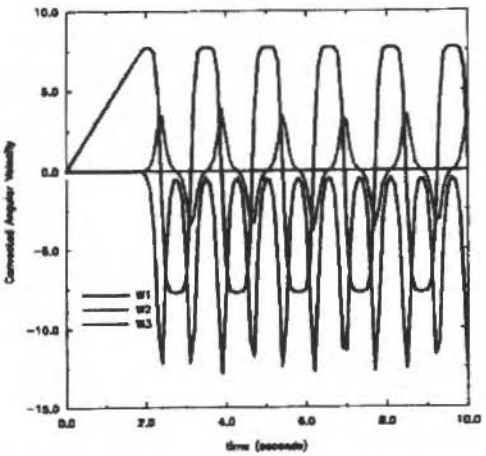


Fig. 15 Convected Angular Velocity

Concluding Remarks

A variational formulation to the rigid's body problem was presented in order to provide a robust formulation, which not only afford a general unified approach, but also is very convenient from the numerical standpoint.

The main question in dealing with large rotations is circumvented by the use of lagrange multipliers, which will imply in the augmentation of the number of degrees of freedom. By the way, this can be avoid in numerical applications using a similar scheme to that used in Rochinha (1990) and Le Tallec et al. (1992), where the lagrange multipliers are computed with a little computational effort.

Finally, a similar approach can be used in the case of flexible links like rods. A first attempt is presented in Campos (1996). So, the present framework seems to be very convenient to be used in the modelling of flexible multibody systems.

References

Antman, S.S. and Kenney, C.S., 1981, "Large Buckled States of Nonlinear Elastic Rods Under Torsion, Thrust and Gravity", Arch. Rat. Mech. Anal. 76, pp. 289-337.

Borri, M., Mello, F. and Atluri, S.N., 1991, "Variational Approaches for Dynamics and Time-Finite Elements: Numerical Studies", Computational Mechanics 7, pp. 49-76.

Campos, A.D. and Rochinha, F.A., 1996, "Dynamics of Rods: Numerical Formulation", Proceedings of Advances in computational Technique for Structural Engineering, Budapest, Hungary, pp. 239-246.

Cardona, A. and Geradian, M., 1989, "Time Integration of the Equations of Motion in Mechanism Analysis", Computers & Structures 33, pp. 801-820.

Downer, J.D., Park, K.C. and Chiou, E., 1992, "Dynamics of Flexible Beams for Multibody Systems: A Computational Procedure", Comp. Meth. Appl. Mech. and Engng. 96, pp. 373-408.

Hughes, P.C., 1986, "Spacecraft Attitude Dynamics", Wiley, New York.

Kolmanovsky, I. and McClamroch, N.H., 1995, "Developments in Nonholonomic Control Problems", IEEE Control Systems Journal 15, pp. 20-36.

Meirovitch, L., 1970, "Methods of Analytical Dynamics", McGraw-Hill, New York.

Rochinha, F.A., 1990, "Modeling and Numerical Simulation of Rods", (in portuguese), PhD Thesis, PUC-Rio, Rio de Janeiro, Brazil.

Sharf, I., 1995, "A Survey of Geometric Stiffening in Multibody Dynamics Formulations", Wave Motion, Intelligent Structures and Nonlinear Mechanics, Vol. 1, pp. 239-279.

Sharf, I., 1996, "Geometrically Nonlinear Beam Element for Dynamics Simulation of Multibody Systems", Int. J. Num. Meth. Engng. 39, pp. 763-786.

Simo, J.C. and Wong, K.K., 1991, "Unconditionally Stable Algorithms for Rigid Body Dynamics that Exactly Preserve Energy and Momentum", Int. J. Num. Meth. Engng. 31, pp. 19-52.

Le Tallec, P. Mani, S.A. and Rochinha, F.A., 1992, "Finite Element Computation of Hyperelastic Rods in Large Displacements", RAIRO - Math. Model. Num. Anal. 26, pp. 325-346.

Vu-Quoc, L. and Simo, J.C., 1988, "Dynamics of Earth-Orbiting Flexible Satellites with Multibody Components", J. Guidance and Control 1, pp. 549-558.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 242-252

ISSN 0100-7386
Printed in Brazil

Control Design for Uncertain Systems Under Fuzzy Disturbance

Y.H. Chen
Georgia Institute of Technology
The George W. Woodruff School of Mechanical Engineering
30332 Atlanta, Georgia USA
yehwa.chen@me.gatech.edu

Abstract

The problem of designing controls for dynamic systems under input disturbance is considered. We consider two possible characterizations of the input disturbance: (1) it is bounded by a crisp number, or (2) it is bounded by a fuzzy number. The control design is purely deterministic. However, the resulting system performance is interpreted differently, depending on the bound information. It may be deterministic or fuzzy (i.e., with a spectrum of possible outcome).
***Keywords**: Uncertain Systems, Fuzzy Set Theory.*

Introduction

Since the inception of fuzzy theory (Zadeh, 1965), the progress in this area has been mainly focusing on the use of fuzzy reasoning for control, estimation, decision making, etc. Fuzzy reasoning (Dubois and Prade, 1991a,b) is indispensable in the sense that such form of reasoning is more aligned with human nature. Furthermore, the information it is based upon (i.e., fuzzy propositions) is hardly of any use to other reasoning mechanism.

On the other hand, as another potential application, fuzzy theory is also a valid tool for describing uncertainty. Exploring uncertainty and determining what is known has been one major task in many scientific disciplines. Once the known portion is clearly identified, human beings may, for example, develop some (e.g., physical) laws which govern various phenomena.

As the other side of the effort, when the known portion can not be completely isolated from the unknown, one may take a more "phenomenological" approach to describe what is known among the whole.

As a very successful endeavor, the *probabilistic* approach takes *the frequency of occurrence* point of view. By this, we mean one may consider an event (such as "head up" in flipping a coin) and then decides how often it occurs. As a typical example, while one does not know the exact outcome of each flip (hence an uncertainty), one knows (or at least assumes subjectively) that the frequency of occurrence of head up, as the number of flip approaches infinity, approaches 0.5 (or in short, the probability of head up is 0.5).

The *fuzzy* approach, on the other hand, takes the *extent of occurrence* point of view. Consider the full occurrence be indexed by 1, one inquires to which extent (which is indexed from 0 to 1) an event occurs. As an example, given that any person with age 75 or higher is old (i.e., the full occurrence of the event "old"), a person with age 67 may be indexed by 0.5, as the extent of being "old".

As a side note, despite that many membership functions, which are used to suggest the index, in fuzzy theory are acquired somehow subjectively (see, e.g., Chapter 10, Klir and Yuan, 1995), one should distinguish fuzzy theory from subjective probability (see, e.g., Kahneman et al., 1982). While the later assumes the probability distribution subjectively, its intention of describing the frequency of occurrence remains. More discussions on the difference (or, as an attack angle, the indifference) between fuzzy theory and probability theory can be found in, e.g., IEEE, 1994.

In this work, we shall attempt to incorporate the fuzzy description of uncertainty into a robust control design framework. The objective is to explore further descriptions of system performance should more information of the uncertainty (in the fuzzy sense) is provided. The basic framework we present is first illustrated by the use of a scalar system. it is then extended to multi-dimensional case.

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors. Agenor de Toledo Fleury and Hans Ingo Weber.

Preliminaries

The following is a brief summary of some basic terms in fuzzy set theory (Wang, 1996).

Universe of discourse. The universe of discourse U is a collection of elements which contain those that are of concern.

Fuzzy set. A fuzzy set D in a universe of discourse U is characterized by a membership function $\mu_D(x)$ that takes value in the interval [0,1] for all x in U.

Crisp set. A special case of fuzzy set as $\mu_D(x)=1$ for all x in U.

Normal fuzzy set. A fuzzy set D in U is normal if there is at least one element x in U such that $\mu_D(x)=1$.

Convex fuzzy set. A fuzzy set D in U is convex if

$$\mu_D(\lambda x_1+(1-\lambda)x_2) \geq min[\mu_D(x_1),\mu_D(x_2)] \tag{1}$$

for all x_1, x_2 in U and all $\lambda \in [0,1]$.

$\overline{\alpha}$ -cut. An α-cut of a fuzzy set D is a crisp set D_α where

$$D_\alpha = \{x \in U \,|\, \mu_D(x) \geq \alpha\} \tag{2}$$

$\overline{\alpha}$ + -cut. An α+ -cut of a fuzzy set D is a crisp set $D_{\alpha+}$ where

$$D_{\alpha+} = \{x \in U \,|\, \mu_D(x) > \alpha\} \tag{3}$$

Support. The support of a fuzzy set D is D_{0+}.

Fuzzy number. Let G be a fuzzy set in **R**, the real number. G is called a fuzzy number if: (i) G is normal, (ii) G is convex, (iii) the support of G is bounded, (iv) all α-cuts are closed intervals in **R**.

Throughout, we shall always assume that the universe of discourse of a fuzzy number to be its 0-cut.

Union. The union of two fuzzy sets D_1 and D_2 in U is a fuzzy set $D_1 \cup D_2$ with

$$\mu_{D_1 \cup D_2}(x) = max\left[\mu_{D_1}(x), \mu_{D_2}(x)\right] \tag{4}$$

Decomposition theorem. Define a fuzzy set $\widetilde{D}_\alpha$ in U with the membership function $\mu_{\widetilde{D}_\alpha} = \alpha I_{A_\alpha}(x)$ where $I_{A_\alpha}(x)=1$ *if* $x \in D_\alpha$ *and* $I_{A_\alpha}(x)=0$ *if* $x \in U - A_\alpha$. Then the fuzzy set D is obtained as

$$D = \bigcup_{\alpha \in [0,1]} \widetilde{D}_\alpha \tag{5}$$

where $\cup$ is the union of the fuzzy sets. The union is taken pair-wisely.

Possibility. Given a fuzzy set D and the proposition "x is in D", the possibility distribution associated with $x = u$, denote by $\pi_X(u)$, is defined to be numerically equal to the grade of membership of u in D; that is,

$$\pi_x(u)=\mu_D(u) \tag{6}$$

for all $u \in U$. In other words, given that x is in D, the possibility that $x = u$ is $\mu_D(u)$.

There are two kinds of interpretation of "belong to" in this paper. An element may belong to a fuzzy set to a certain degree (hence such a "belong to" is in a *fuzzy* sense). However, as an exception, whenever we say an element belongs to a (crisp) interval of the real number, it is always meant in the *classical* sense.

Scalar Case

Consider the following class of scalar systems

$$\dot{x}(t)=ax(t)+bu(t)+v, \qquad x(t_0)=x_0 \tag{7}$$

where $t \in \mathbf{R}$ is the "time" (or more precisely, the independent variable), $x(t) \in \mathbf{R}$ is the state, $u(t) \in \mathbf{R}$ is the control, $v \in \mathbf{R}$ is the (constant) input disturbance. Furthermore, a and b are constants. The input disturbance and initial condition are unknown. However, we consider the following two possible characterizations.

Assumption 1(a). *v is unknown but bounded.* There exists a known constant $\hat{V}(\geq 0)$ such that

$$|v| \leq \hat{V} \tag{8}$$

Remark. This is often described as the worst case description. Only the *maximum* possible bound of v is known.

Assumption 1(b). *x_0 is bounded.* There exists a known constant $\eta(\geq 0)$ such that

$$|x_0| \leq \eta \tag{9}$$

Assumption 2(a). *v is a fuzzy disturbance.* There exists a known fuzzy number V such that v is in V with $\mu_V(v)$. The universe of discourse of V is its 0-cut.

Assumption 2(b). *x_0 is a fuzzy initial condition.* There exists a known fuzzy number X such that x_0 is in X with $\mu_X(x_0)$. The universe of discourse of X is its 0-cut.

Remark. Assumption 2 provides more information regarding the possible bound: The bound is close to a value ξ which is such that $\mu_V(\xi) = 1$. The degree of closeness is described by the membership function.

We now consider the control u to be

$$u(t)=-kx(t) \tag{10}$$

where k is a constant such that $\bar{a} := a - bk < 0$. Then the solution of the (closed-loop) system is given by, for all $t \geq t_0$,

$$\begin{aligned} x(t) &= e^{\bar{a}(t-t_0)}x_0 + \int_{t_0}^{t} e^{\bar{a}(t-\tau)}v\,d\tau \\ &= e^{\bar{a}(t-t_0)}x_0 + \frac{1}{\bar{a}}(e^{\bar{a}(t-t_0)}-1)v \end{aligned} \tag{11}$$

Note that $(e^{\bar{a}(t-t_0)}-1)/\bar{a} \geq 0$ for all $t \geq t_0$ since $\bar{a}<0$. If v and x_0 are described by Assumption 1, then we have, based on (11),

$$|x(t)| \leq e^{\bar{a}(t-t_0)}|x_0|+\frac{1}{\bar{a}}(e^{\bar{a}(t-t_0)}-1)|v|$$

$$\leq \beta_1(t)_\eta+\beta_2(t)\hat{V} \tag{12}$$

where

$$\beta_1(t)=e^{\bar{a}(t-t_0)}, \tag{13}$$

$$\beta_2(t)=\frac{1}{\bar{a}}(e^{\bar{a}(t-t_0)}-1) \tag{14}$$

Notice that $\beta_1(t) > 0$ and $\beta_2(t) \geq 0$ for all $t \geq t_0$. This enables us to conclude, for any prescribed $\underline{d} > -\hat{V}/\bar{a}$ and any initial condition with $\eta > \underline{d}$, there is a finite time $T(\eta,\underline{d})$ with

$$T(\eta,\underline{d})=\frac{1}{\bar{a}}ln\left(\frac{\underline{d}+\frac{\hat{V}}{\bar{a}}}{\eta+\frac{\hat{V}}{\bar{a}}}\right) \tag{15}$$

such that $|x(t)| \leq \underline{d}$ for all $t \geq t_0+T(\eta,\underline{d})$. This performance is called the uniform ultimate boundedness (Chen and Leitmann, 1987). We summarize the performance as follows.

Lemma 1. Consider the system (7) under the control (10). Subject to Assumption 1, the state $x(t)$ is uniformly ultimately bounded.

Remark. Since one can choose k to make $\bar{a}$ to be arbitrarily negative, $\underline{d}$, the size of the region which $x(t)$ will eventually enter, can be made arbitrarily small. However, for any $\underline{d}$, with $|x(t)| \leq \underline{d}$, it does not further indicate which area inside this region $x(t)$ may reside. This is a worst case analysis only.

We now consider that further information of v and x_0 are known so that they can be cast into Assumption 2.

Since V is fuzzy set, for each $\alpha \in [0,1]$, its α-cut corresponds to a (crisp) interval $[\underline{v}_\alpha, \bar{v}_\alpha]$. Similarly, the α-cut of X is a (crisp) interval $[\underline{x}_{0\alpha}, \bar{x}_{0\alpha}]$.

By fuzzy arithmetics (Wang, 1996), for each $\alpha \in [0,1]$, we consider, in view of the right hand side of (11), the (crisp) interval

$$\left[\beta_1(t)\underline{x}_{0\alpha}+\beta_2(t)\underline{v}_\alpha, \beta_1(t)\bar{x}_{0\alpha}+\beta_2(t)\bar{v}_\alpha\right]=: D_\alpha(t) \tag{16}$$

That (11) holds for all $t \geq t_0$ enables us to draw the following two interpretations of the result.

The result may be interpreted by the use of *possibility theory* (Zadeh, 1978) as follows. At time t, for given $\alpha \in [0,1]$, there corresponds to a region D_α which is closed and bounded. The region D_α can be considered an α-cut of a membership function $\mu_D(x) \in [0,1]$. This is through the use of the decomposition theorem. This in turn defines a fuzzy set D, which is in fact also a fuzzy number, whose universe of discourse is D_0. Given that $x(t) \in D$, the possibility of $x = \zeta$ at time t is then equal to $\mu_D(\zeta)$.

An alternative interpretation is via the use of *confidence index* as follows: Consider the fuzzy set D as mentioned above. For each α, the confidence index is defined to be 1-α. For given $\alpha \in [0,1]$, the value of x at time t lies in the internal $[\beta_1(t)\underline{x}_{0\alpha} + \beta_2(t)\underline{v}_{-\alpha}, \beta_1(t)\bar{x}_{0\alpha} + \beta_2(t)\bar{v}_\alpha]$ (this is interpreted in the classical sense) is assured by the confidence index 1-α. For example, as $\alpha = 0$, the interval is in fact the universe of discourse of the fuzzy set D. The confidence index that x is in this interval is 1. In other words, that x lying within this interval is assured. In the special case that the confidence index of x in an interval is equal to 1 for all $t \geq t_0$, the state x is called *uniformly bounded* (Chen and Leitmann, 1987).

Our purpose, from the control point of view, is to prescribe a region in which the state will lie. This prompts the following analysis. We first prescribe an interval $[\underline{d}_\alpha, d_\alpha]$, $\alpha \in [0,1]$, and a finite time T. The problem is to choose the control gain k such that

$$x(t) \in \left[\underline{d}_\alpha, \bar{d}_\alpha\right] \tag{17}$$

(this is interpreted in the classical sense) for all $\alpha \in [0,1]$ and $t \geq t_0 + T$.

We shall proceed as follows. First, based on (17), one intends to have, for all $t \geq t_0 + T$

$$e^{\bar{a}(t-t_0)}\underline{x}_{0\alpha} + \frac{1}{\bar{a}}(e^{\bar{a}(t-t_0)} - 1)\underline{v}_\alpha \geq \underline{d}_\alpha \tag{18}$$

This means that

$$e^{\bar{a}(t-t_0)}(\underline{x}_{0\alpha} + \frac{\underline{v}_\alpha}{\bar{a}}) \geq \underline{d}_\alpha + \frac{\underline{v}_\alpha}{\bar{a}} \tag{19}$$

Second, one also intends to have, for all $t \geq t_0 + T$,

$$e^{\bar{a}(t-t_0)}\bar{x}_{0\alpha} + \frac{1}{\bar{a}}(e^{\bar{a}(t-t_0)} - 1)\bar{v}_\alpha \leq \bar{d}_\alpha \tag{20}$$

or

$$e^{\bar{a}(t-t_0)}(\bar{x}_{0\alpha} + \frac{\bar{v}_\alpha}{\bar{a}}) \leq \bar{d}_\alpha + \frac{\bar{v}_\alpha}{\bar{a}} \tag{21}$$

We first outline the following two results.

Theorem 1. Consider the system (7) under the control (10). Subject to Assumption 2, there exists a k with $\bar{a} < 0$ such that the following four cases hold.

(i) If $\underline{x}_{0\alpha} + \underline{v}_\alpha / \bar{a} \geq 0, \underline{d}_\alpha + \underline{v}_\alpha / \bar{a} \leq 0$, then $x(t) \geq \underline{d}_\alpha$ for all $t \geq t_0 + \underline{T}$ where $\underline{T}$ is any non-negative constant, $\alpha \in [0,1]$.

(ii) If $\underline{d}_\alpha + \underline{v}_\alpha / \bar{a} \leq \underline{x}_{0\alpha} + \underline{v}_\alpha / \bar{a} < 0$, then $x(t) \geq \underline{d}_\alpha$ for all $t \geq t_0 + \underline{T}$ where $\underline{T}$ is any non-negative constant, $\alpha \in [0,1]$.

(iii) If $\underline{x}_{0\alpha} + \underline{v}_\alpha / \bar{a} < \underline{d}_\alpha + \underline{v}_\alpha / \bar{a} < 0$, then $x(t) \geq \underline{d}_\alpha$ for all $t \geq t_0 + \underline{T}$ where

$$\underline{T} = \sup_{\alpha \in [0,1]} \frac{1}{\bar{a}} ln \left(\frac{\underline{d}_\alpha + \frac{\underline{v}_\alpha}{\bar{a}}}{\underline{x}_{0\alpha} + \frac{\underline{v}_\alpha}{\bar{a}}} \right) \tag{22}$$

$$\alpha \in [0,1]$$

(iv) For any other cases of $\underline{x}_{0\alpha} + \underline{v}_\alpha / \bar{a}$ and $\underline{d}_\alpha + \underline{v}_\alpha / \bar{a}$, there does not exist a $\underline{T} \geq 0$ such that $x(t) \geq \underline{d}_\alpha$ for all $t \geq t_0 + \underline{T}, \alpha \in [0,1]$.

Remark. Cases (i) and (ii) are trivial cases. The result in case (iii) can be used in a constructive way: By prescribing $[\underline{d}_\alpha, d_\alpha]$, and $\underline{T}$, one finds a suitable k, which is in fact non-unique, to achieve the task. Case (iv) shows the limit of the control design.

Theorem 2. Consider the system (7) under the control (10). Subject to Assumption 2, there exists a k with $\bar{a} < 0$ such that the following four cases hold.

(i) If $\bar{x}_{0\alpha} + \bar{v}_\alpha / \bar{a} \leq 0, \bar{d}_\alpha + \bar{v}_\alpha / \bar{a} \geq 0$, then $x(t) \leq \bar{d}_\alpha$ for all $t \geq t_0 + \bar{T}$ where $\bar{T}$ is any non-negative constant, $\alpha \in [0,1]$.

(ii) If $\bar{d}_\alpha + \bar{v}_\alpha / \bar{a} \leq \bar{x}_{0\alpha} + \bar{v}_\alpha / \bar{a} > 0$, then $x(t) \leq \bar{d}_\alpha$ for all $t \geq t_0 + \bar{T}$ where $\bar{T}$ is any non-negative constant, $\alpha \in [0,1]$.

(iii) If $\bar{x}_{0\alpha} + \bar{v}_\alpha / \bar{a} > \bar{d}_\alpha + \bar{v}_\alpha / \bar{a} > 0$, then $x(t) \leq \bar{d}_\alpha$ for all $t \geq t_0 + \bar{T}$ where

$$\bar{T} = \sup_{\alpha \in [0,1]} \frac{1}{\bar{a}} ln \left(\frac{\bar{d}_\alpha + \frac{\bar{v}_\alpha}{\bar{a}}}{\bar{x}_{0\alpha} + \frac{\bar{v}_\alpha}{\bar{a}}} \right) \tag{23}$$

$$\alpha \in [0,1]$$

(iv) For any other cases of $\bar{x}_{0\alpha} + \bar{v}_\alpha / \bar{a}$ and $\bar{d}_\alpha + \bar{v}_\alpha / \bar{a}$, there does not exist a $\bar{T} \geq 0$ such that $x(t) \geq \bar{d}_\alpha$ for all $t \geq t_0 + \bar{T}, \alpha \in [0,1]$.

Similar comments as the last remark can be made on this theorem. Finally, we can summarize the results as following.

Lemma 2. Consider the system (7) under the control (10) with $\bar{a} < 0$ and subject to Assumption 2. Consider the prescribed interval $[\underline{d}_\alpha, \bar{d}_\alpha]$ for each $\alpha \in [0,1]$ and the resulting membership function $\mu_D(x)$ for the fuzzy number D. For given $\alpha \in [0,1]$, the value of x at any time $t \geq t_0 + \hat{T}$, where

$$\hat{T} = max\left\{\underline{T}, \bar{T}\right\} \tag{24}$$

lies in the interval is assured by the confidence index 1-α.

Remark. The result renders uniform boundedness and uniform ultimate boundedness (Chen and Leitmann, 1987) a special case (as $\alpha = 0$).

Remark. An interpretation using the possibility theory can also be drawn.

Multi-dimensional Case

The previous analysis considered two possible ways of characterizing the input disturbance. The first is that the input disturbance is bounded and the (crisp) bound is known. The second is that the input disturbance is within a fuzzy set. In practice, it is in fact more desirable to combine these two characterizations. This will be treated in this section. We also consider the multi-dimensional case. Consider the following uncertain system

$$\dot{x}(t) = Ax(t) + B(u(t) + v(t)), \qquad x(t_0) = x_0 \tag{25}$$

where $t \in \mathbf{R}$, $x(t) \in \mathbf{R}^n$ is the state, $u(t) \in \mathbf{R}^m$ is the control, $v(t) \in \mathbf{R}^m$ is the (unknown) time-varying input disturbance, A, B are constant matrices. The function $v(\cdot)$ is Lebesgue measurable. The following assumption are made.

Assumption 3. The pair (A,B) is stabilizable.

Assumption 4. There is a (not necessarily known) scalar $\nu \geq 0$ such that

$$\max_{t \in R} \left\| v(t) \right\| \leq \nu \tag{26}$$

The scalar ν is in a fuzzy number N which is prescribed by a membership function $\mu_N(\nu) \in [0,1]$.

Remark. Assumption 4 is, in its essence, a fuzzy description of the knowledge of the possible bound of $v(t)$. This can be viewed as a combination of Assumptions 1 and 2 but for multi-dimensional case.

The task is to choose the control u such that the state belongs to a region around $x = 0$ after a finite time and remains there thereafter.

First choose a gain matrix K such that $\bar{A} := A - BK$ is Hurwitz. This is feasible if (A,B) is stabilizable. For any $n \times n$ matrix $Q > 0$, one solves for the unique solution $P > 0$, which is an $n \times n$ matrix, from the following Lyapunov equation

$$\bar{A}^T P + P\bar{A} + Q = 0 \tag{27}$$

We propose the control u as follow:

$$u(t) = -Kx(t) - \gamma B^T P x(t) \tag{28}$$

where $\gamma > 0$ is a scalar and is to be specified later.

Theorem 4. Consider the system (25) subject to Assumptions 3 and 4. Suppose that the control (28) is applied. Consider a prescribed interval $[0, \bar{d}_\alpha]$ for each $[\alpha \in 0,1]$. This is chosen such that the resulting fuzzy set, which is obtained via the decomposition theorem, can be a fuzzy number. Then there exists a $\gamma > 0$ and a finite time $\widetilde{T} \geq 0$ such that the value of $\|x(t)\|$ at any time $t \geq t_0 + \widetilde{T}$ lies in the interval $[0, \bar{d}_\alpha]$ is assured by the confidence index 1 - α.

Proof. We prove this via the Lyapunov minimax approach (Corless, 1993; Leitmann, 1993). Consider the Lyapunov function candidate

$$V(x) = x^T P x \tag{29}$$

For any admissible $v(\cdot)$, its time derivative along the trajectory of the controlled system of (25) is given by (arguments are sometimes omitted when no confusions are likely to arise)

$$\dot{V} = 2x^T P\left[Ax + B(-Kx - \gamma B^T P x) + Bv\right]$$

$$= x^T(P\bar{A} + \bar{A}^T P)x - 2\gamma \left\| B^T P x \right\|^2 + 2x^T P B v \tag{30}$$

By the use of the Lyapunov equation (29) and since

$$x^T P B v \leq \left\| B^T P x \right\| v \tag{31}$$

$$\dot{V} \leq -x^T Q x - 2\gamma \left\| B^T P x \right\|^2 + 2v \left\| B^T P x \right\| \tag{32}$$

By the Rayleigh's principle (Franklin, 1968),

$$x^T Q x \geq \lambda_{min}(Q) \| x \|^2 \tag{33}$$

and hence

$$-x^T Q x \leq -\lambda_{min}(Q) \| x \|^2 \tag{34}$$

Furthermore,

$$-\gamma \| B^T P x \|^2 + \nu \| B^T P x \| \le \frac{\nu^2}{4\gamma} =: \delta \tag{35}$$

With these back into (32), we have

$$\dot{V} \le -\lambda_{min}(Q) \| x \|^2 + \delta \tag{36}$$

Hence $\dot{V}$ is negative definite for all $\| x \|$ such that

$$\| x \| > \sqrt{\frac{\delta}{\lambda_{min}(Q)}} =: \underline{d} \tag{37}$$

By Chen and Leitmann (1987), the system is uniformly ultimately bounded in the sense that with $\| x_0 \| \le r$,

$$\| x(t) \| \le \bar{d} \quad \text{for all } t \ge t_0 + \tilde{T}$$

with any

$$\bar{d} > \delta \sqrt{\frac{\lambda_{max}(P)}{\lambda_{min}(P)}} \tag{38}$$

and a corresponding finite time

$$\tilde{T} = 0 \quad \text{if } r \le \bar{d} \sqrt{\frac{\lambda_{min}(P)}{\lambda_{max}(P)}} \tag{39}$$

$$\tilde{T} = \frac{\lambda_{max}(P) r^2 - \lambda_{min}(P) \bar{R}^2}{\lambda_{min}(Q) \bar{R}^2 - \delta} \quad \text{otherwise,} \tag{40}$$

where

$$\bar{R} = \bar{d} \sqrt{\frac{\lambda_{min}(P)}{\lambda_{max}(P)}} \tag{41}$$

The performance stated in the theorem then follows.

Remark. For simplicity, we do not consider the fuzziness regarding the bound of initial condition in this theorem. Hence $\| x_0 \| \le r$ is stated in the classical sense.

Remark. The proof in fact also suggests a way to select γ: For the prescribed $\bar{d}_\alpha > 0$ (this is the designer's discretion), one chooses any $\underline{d}_\alpha < \bar{d}_\alpha$ and the corresponding δ according to (37). The γ is then selected via (35) for all $\alpha \in [0,1]$. In a sense, the choice of γ is similar to some other work (e.g., Barmish et al., 1983). However, the interpretation of the performance (via the confidence index) includes the previous performance (i.e., uniform ultimate boundedness) as a special case. This is mainly due to that more information of the disturbance bound is used.

Remark. Both the disturbance bound characterization and the performance are described in a fuzzy sense. This renders some early work, which described them via the worst case scenario, a special case.

Remark. The current fuzzy description are very much aligned with practice. The input disturbance bound is often obtained via experimental data and analyzed by the engineer. One may judge the bound to be, for instance, "close to" a (crisp) value or "very close to" a (crisp) value. These are standard fuzzy (linguistic) terms. The performance is also often judged by the engineer depending on one's need: One may choose a (crisp) set point and intend to have the performance to be "close to" or "very close to" it, after a finite time. In addition, the engineer may also impose a hard bound on the performance, which must be met (hence this corresponds to uniform ultimate boundedness). These can be all addressed by the current framework.

Conclusions

The incorporation of uncertainty, which is described in a fuzzy sense, into a robust control framework is introduced. This is believed to the first attempt for such a merge. Previously the control design was only based on the maximum possible bound of uncertainty. The system performance was described in a worst case scenario. The current extension into fuzzy domain enables one to draw further information of the systems performance should further information of the uncertainty is available. As to the prescription of the desirable performance, it is often the designer's discretion. Since in practice it is in fact more realistic to prescribe the performance in a fuzzy sense (such as "close", "very close"), the current framework fits in well with both the need (the performance) and the given (uncertainty).

References

Barmish, B.R., Corless, M. and Leitmann, 1993, "A New Class of Stabilizing Controllers for Uncertain Dynamical Systems", SIAM Journal of Control and Optimization, Vol. 21, pp. 246-255.

Chen, Y.H. and Leitmann, G., 1987, "Robustness of Uncertainty Systems in the Absence of Matching Assumptions", International Journal of Control, Vol. 45, pp. 1527-1542.

Corless, M., 1993, "Control of Uncertain Nonlinear Systems", Journal of Dynamic Systems, Measurement, and Control, Vol. 115, pp. 362-372.

Dubois, D. and Prade, H. 1991a, "Fuzzy sets in Approximate Reasoning, Part 1: Inference with Possibility Distributions", Fuzzy Sets and Systems, Vol. 40, pp. 143-201.

Dubois, D. and Prade, H. 1991b, "Fuzzy Sets in Approximate Reasoning, Part 2: Logical Approaches", Fuzzy Sets and Systems, Vol. 40, pp. 202-244.

Franklin, J.N., 1968, Matrix Theory, Prentice-Hall, Englewood Cliffs, NJ.

IEEE 1994, "Special Issue - Fuzziness vs. Probability - the N-th Round", IEEE Transactions on Fuzzy Systems, Vol. 2, pp. 1-45.

Kahneman, D., Slovic, P. and Tversky, A., 1982, Judgment Under Uncertainty: Heuristics and Biases, Cambridge University Press, Cambridge, NY.

Klir, G.J. and Yuan, B., 1995, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall, Englewood Cliffs, N.J.

Leitmann, G., 1993, "On One Approach to the Control of Uncertain Systems", Journal of Dynamic Systems, Measurement and Control, Vol. 115, pp. 373-380.

Wang, L.X., 1996, A Course in Fuzzy Systems and Control, Prentice-Hall, Upper Saddle River, NJ.
Zadeh, L.A., 1965, "Fuzzy Sets", Information and Control, Vol. 8, pp. 338-353.
Zadeh, L.A., 1978, "Fuzzy Sets as a Basis for a Theory of Possibility", Fuzzy Sets and Systems, Vol. 1, pp. 3-28.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 253-261

ISSN 0100-7386
Printed in Brazil

Computation of Lyapunov Exponents Using Householder Factorization

Hubertus F. von Bremen
vonbrem@aludra.usc.edu

Firdaus E. Udwadia
firdwadia@altinak.usc.edu
University of Southern California
Department of Mechanical Engineering
CA 90089-1453 USA

Wlodek Proskurowski
University of Southern California
Department of Mathematics
CA 90089-1113 USA

Abstract

An efficient and numerically stable method to determine all the Lyapunov characteristic exponents of a dynamical system is presented. The method is compared with known methods in terms of efficiency, and the accuracy and stability of the methods are tested by numerical experiments.
Keywords: *Nonlinear Dynamical Systems, Lyapunov Exponents, Householder Factorization.*

Introduction

The Lyapunov characteristic exponents (LCE's) are important for the understanding of the dynamics of nonlinear dynamical systems. A solid analytical basis for the computation of the LCE's was given by Benettin et al. (1980) in their two part paper. Benettin et al. (1978) are the first to propose a Gram-Schmidt orthogonalization type procedure to compute the LCE's.

Geist et al. (1990) made a through comparison of several methods for computing LCE's and also presented the main ideas published in the previous decade (Shimada and Nagashima (1979), Benettin et al. (1980), Wolf et al. (1985), Eackmann and Ruelle (1985). For additional more recent references see Barna and Tsuda (1993), and Dieci and Van Vleck (1995).

For discrete systems, the computation of the LCE's involves the factorization of a matrix into the product of an orthogonal matrix Q and an upper triangular matrix R (a QR-factorization). This factorization can be obtained by using the Gram-Schmidt (GS) orthogonalization, the Modified GS (MGS) orthogonalization or the Householder orthogonal factorization (HQR). Among these methods, the GS is known to be numerically unstable, since the orthogonal matrix Q may deviate greatly from orthogonality due to accumulation of roundoff errors. This is why MGS is preferred, see Parker and Chua (1989), Dieci and Van Vleck (1995). On the other hand, the HQR method is known to be backward stable, see Wilkinson (1965).

For an n by n matrix the asymptotic (for large n) cost of MGS (or GS) factorization is $2n^3$ flops (a flop is a floating point addition or multiplication, Golub and Van Loan (1993), while the HQR-factorization requires $4/3n^3$ flops to compute the upper triangular R, and an additional $2n^3$ flops if also the orthogonal Q is required. In computing the LCE's one also needs to multiply each of the Q matrices (computed by the QR-factorization) by the consecutive matrix representation of the tangent map. The usual use of the HQR method (see Geist et al. (1990) in the factorization and multiplication of Q with the tangent map is computationally more expensive than the MGS method. In this paper we show how to organize computation of the LCE's using HQR-factorization in such a manner that computational savings are obtained over the MGS based method.

The efficiency of the different methods is addressed in the paper. The accuracy of the methods is also illustrated by the use of numerical experiments.

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

Computation of LCE's using the QR-factorization

In smooth dynamical systems, we usually deal with maps of the form $x^t = T^t x$ (here x often belongs to a suitable compact connected manifold M, and the map is from a (finite) n-dimensional space to an n-dimensional space) with $T^t = T \circ T^{t-1}$. The parameter t denotes a nonnegative integer or a real number. The sequence of tangent maps dT_x^t is obtained through iterations by considering $dT_x^t = dT_{T^t x} \circ dT_x^{t-1}$. We shall consider the matrix representation of these operators in the standard bases (in fact, any orthonormal basis set would be sufficient). The LCE's can then be obtained from the QR-factorization of the product of the matrix representations of these tangent maps which are determined at the appropriate points x of M. For brevity we shall denote the matrix representation of the tangent map (evaluated at the appropriate point x of the manifold) corresponding to $t = i$ by J_i. To determine an approximation to all the LCE's, one then needs the QR-factorization of the matrix product $J_m J_{m-1} \ldots J_1$. This decomposition can be done sequentially as follows. Starting with $Q_0 = I$, we have,

$$\begin{aligned} qr\left[J_m J_{m-1} \ldots J_1\right] &= qr\left[J_m J_{m-1} \ldots J_2\left(J_1 Q_0\right)\right] = qr\left[J_m J_{m-1} \ldots J_3\left(J_2 Q_1\right)\right]\left[R_1\right] \\ &= qr\left[J_m J_{m-1} \ldots \left(J_3 Q_2\right)\right]\left[R_2 R_1\right] = \ldots = \\ &= qr\left[J_m J_{m-1} \ldots \left(J_i Q_{i-1}\right)\right]\left[R_{i-1} R_{i-2} \ldots R_2 R_1\right] = \ldots = \\ &= Q_m\left[R_m \ldots R_2 R_1\right] = Q_m R \end{aligned} \tag{1}$$

Here we sequentially use the QR-factorization, and qr[.] denotes the QR-factorization process. Starting with J_1, at each step i in the above sequence, we perform a premultiplication $B_i = J_i Q_{i-1}$ followed by a QR-factorization of $B_i = J_i Q_{i-1} = Q_i R_i$, $i = 1, 2, \ldots, m$. The matrix R is the product of the matrices $R_m \ldots R_2 R_1$ obtained in this sequential manner. Furthermore, each of the diagonal elements of R, is simply the product of the corresponding diagonal elements of all the R_i's. Hence, approximations to the n LCE's are then obtained as: $X_k = \frac{1}{m} \sum_{i=1}^{m} ln\left|R_i(k,k)\right|$, $k = 1, 2, \ldots, n$. The computation can be presented as:

Algorithm for computing all the LCE's of a dynamical system

Initialization

Initialize Q to be the n by n Identity Matrix

Initialize *LCEvector* to be a zero n-vector

for $i = 1$ *to m_iterations*

$B = J_i Q$

Compute the QR factorization of B $(QR = B)$

$LCEvector = LCEvector + log\left(diag\left(|R|\right)\right)$

end

$LCEvector = LCEvector/m\_iterations.$

The approximate values of the n LCE's after *m_iterations* are then given by the components of the n-vector denoted *LCEvector*. Here the successive maps J_i at each iteration are assumed to be known.

There are several ways of computing the QR-factorization (which is required at each step i) indicated in the loop above, the commonest ones being the Gram-Schmidt QR (GS), the modified Gram-Schmidt (MGS), and the Householder QR (HQR). The choice of method must be based upon factors such as accuracy of the procedure, storage requirements and simplicity of implementations. If the aim is ease of implementation, one could make use of computing environments which have built-in QR-factorizations. For example, MATLAB has a reliable and efficient QR-factorization using Householder reflectors.

The Householder QR-factorization is known to be backward stable (Wilkinson, 1965) (with regard to roundoff errors). Yet its direct application is (asymptotically, for large n) computationally more expensive than the GS (or the MGS) approach. However, efficiencies in the Householder based QR-factorization in terms of both computation and storage can be achieved because: (1) we need to compute and store only the diagonal entries of the matrix R, and (2) the reflector (Householder) matrices which constitute Q can be sequentially assembled resulting in considerable efficiencies in the computation of the action of Q on the succeeding map J_{i+1}. In what follows we show that by modifications of the standard Householder QR-factorization we obtain a method that while being more computationally efficient that the GS or the MGS approaches seems also more stable with regard to roundoff errors.

Householder QR-based (HQRB) Algorithm for the Computation of LCE's

Consider the QR-factorization of the above mentioned algorithm at the iteration index i. The Householder QR factorization works along the following lines (for details see Dahlquist and Björck (1997). Given an n by n matrix B, one sequentially determines the matrices

$$B^{(s+1)} = H^{(s)} B^{(s)},\ s = 1, 2, \ldots, n-1;\ B^{(1)} \equiv B \tag{2}$$

The Householder reflector matrices $H^{(s)}$ have the structure

$$H^{(s)} = I_n - w^{(s)} \left[w^{(s)} \right]^T \tag{3}$$

where the first $(s - 1)$ elements of the n-vector $w^{(s)}$ are all zero. The matrix R of the QR factorization of B is then obtained as

$$H^{(n-1)} = H^{(2)} H^{(1)} B = R,$$

and the matrix Q is given by

$$Q = H^{(1)} H^{(2)} \ldots H^{(n-1)} \tag{4}$$

We note that this factorization is required to be done at each iteration i described in the aforementioned algorithm. At the next iteration (with index $i + 1$) the matrix B is replaced by the matrix (we suppress the subscript in J_{i+1} and write J for simplicity)

$$JQ = \{JH^{(1)}H^{(2)}\ldots H^{(s)}\}H^{(s+1)}\ldots H^{(n-1)} = J^{(s)}H^{(s+1)}\ldots H^{(n-1)} \tag{5}$$

We now make two important observations:

1.) The LCE's are obtained only through the determination of the diagonal elements of the R_i, $i = 1, 2, \ldots, m$ in equation (1) and therefore one needs to judiciously compute the elements in each of these upper triangular matrices; storage of only the diagonal elements is called for. Using equation (2) with $s = 1$, we get a matrix of the form:

$$B^2 = H^1 B^1 = \begin{bmatrix} \otimes & * & \ldots & * \\ 0 & \otimes & \ldots & \otimes \\ 0 & \otimes & \ldots & \otimes \end{bmatrix} \tag{6}$$

Only the first element in the first column of B^2 needs to be computed, since the rest of the elements are zero. Also in the first row, the only element that is needed is the first element (this is the first diagonal element of R). This is because the other elements denoted with * have no influence in the computation of the rest of the diagonal elements of R. A similar pattern is observed when we apply the rest of the reflectors, as in equation (2).

2.) The action of the premultiplication of Q by J is given by equation (5) so that one does not need to compute Q explicitly. The matrix product $J^{(s)}H^{(s+1)}$ can be written as

$$J^{(s)}H^{(s+1)} = \begin{bmatrix} x^{(s)} & y^{(n-s)} \\ z^{(n-s)} & J^{(n-s)} \end{bmatrix} \begin{bmatrix} I_s & 0 \\ 0 & \tilde{H}_{n-s} \end{bmatrix} \tag{7}$$

where the matrix $x^{(s)}$ is s, the matrix $y^{(n-s)}$ is s by $(n-s)$, the matrix $z^{(n-s)}$ is $(n-s)$ by s $J^{(n-s)}$ is $(n-s)$ by $(n-s)$. This simplifies to

$$J^{(s)}H^{(s+1)} = \begin{bmatrix} x^{(s)} & y^{(n-s)} \\ z^{(n-s)} & J^{(n-s)} \end{bmatrix} \begin{bmatrix} I_s & 0 \\ 0 & I_{n-s} - w^{(n-s)}\left(w^{(n-s)}\right)^T \end{bmatrix}$$

$$= \begin{bmatrix} x^{(s)} & y^{(n-s)} - \left(y^{(n-s)}w^{(n-s)}\right)\left(w^{(n-s)}\right)^T \\ z^{(n-s)} & J^{(n-s)} - \left(J^{(n-s)}w^{(n-s)}\right)\left(w^{(n-s)}\right)^T \end{bmatrix} \tag{8}$$

Thus the first s columns of the matrix remain unchanged after the multiplication, and therefore do not need to be computed. The pseudo-code using the Householder based method for computing the LCE's having an n by n tangent map can be expressed as follows[1].

Pseudo-Code of the Householder QR Based Method (HQRB) for the Computation of all LCE's

Initializations:

Initialize J to be the first tangent map

---

[1] A computer code implementing the pseudo code can be obtained from the authors.

Initialize *Jplus1* to be the next tangent map

Initialize *LCEvector* to be a zero vector

for $i = 1$ to *m_iterations*

QR-Factorization Part

for $k = 1$ to $n - 1$

Computation of the reflectors

The *k-th* reflector is stored in the *k-th* column of *J*. The diagonal elements of *R* are stored in the vector *r*. The computations are as follows:

$$sigma = \sqrt{\sum_{s=k}^{n} J(s,k)^2},$$

$$gamma = sigma\left(sigma + \left|J(k,k)\right|\right)$$

$$r(k) = -sign(J(k,k))\, sigma$$

$$J(k,k) = J(k,k) - r(k)$$

Computation of reflectors with *J*

for $j = k + 1$ to n

$$beta = \left(\sum_{s=k}^{n} J(s,k)\, J(s,j)\right) / gamma)$$

for $s = k + 1$ to n

$J(s,j) = J(s,j) - J(s,k)\ beta$

end (s)

end (j)

Computation of action of Jplus1 on 0

for $j = 1$ to n

$$beta = \left(\sum_{s=k}^{n} Jplus1(j,s)\, J(s,k)\right) / gamma$$

for $s = k$ to n

$Jplus1(j,s) = J\,plus1(j,s) - J(s,k)\ beta$

end (s)

end (j)

end (k)

$r(n) = J(n,n)$

Set $J = Jplus1$

Set $Jplus1$ to be the next tangent map

For $q = 1$ to n

$LCEvector(q) = LCEvector(q) + \log(|r(q)|)$

end (q)

end (i)

$LCEvector = LCEvector / m\_iterations$

Computational Efficiency Comparison

Each iteration in the computation of the LCE's involves one QR-factorization followed by the action of multiplying the succeeding tangent map by the orthogonal matrix Q, thus an operation count for each of these two steps would give a measure of efficiency of the method used. We follow the GS and MGS algorithms given by Golub and Van Loan (1993), and the HQR as given by Dahlquist and Björck (1997). The HQR presented computes R and Q explicitly, but in the computation of Q the explicit product of the reflector matrices (as in equation (4)) is avoided, see also Geist (1990). The HQRB algorithm is the one described in this paper.

The main difference in efficiency of the various methods comes from the way the QR-factorization part of the method is performed. The operation count for the action of multiplying the succeeding tangent map by the orthogonal matrix Q is roughly the same for all methods. The ratios of the asymptotic constant (for large n) in the operation count for the QR-factorization and for the action of multiplying the succeeding tangent map by the orthogonal matrix Q are 6 : 8 : 5 for the GS (and MGS), HQR, and HQRB, respectively. This means that for large enough values of n, the savings by using the HQRB versus the MGS (the more efficient of the remaining methods) in the number of operations are more than 10% (up to about 16% asymptotically). For small values of n, a similar ordering of the methods based on the operation count is observed. Figure 1 shows the number of operations versus system size for the various methods for $n < 10$.

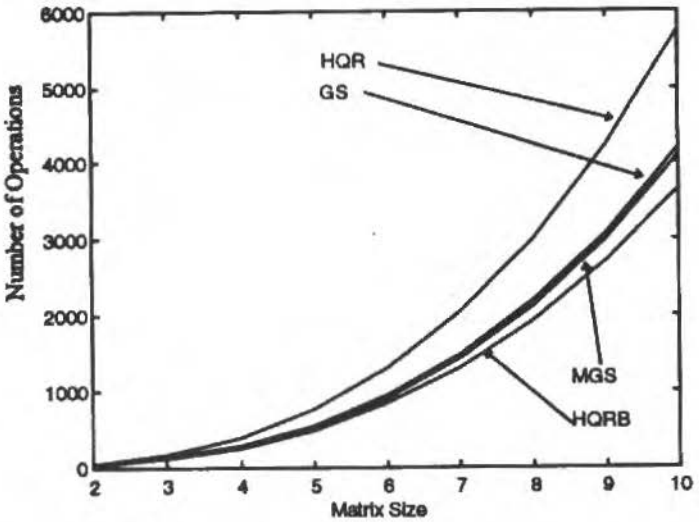


Fig. 1 Number of Operations versus Matrix Size for the GS, MGS, HQR and HQRB

Accuracy Comparison

The following constant map is used to compare the accuracy of the GS, the MGS and the HQRB methods.

$$A=\begin{bmatrix} \frac{110+11\mu}{10} & 1 & 0 & 0 \\ -\frac{100+121\mu}{10} & 0 & 1 & 0 \\ \frac{(110+11\mu)\mu}{10} & 0 & 0 & 1 \\ -\mu^2 & 0 & 0 & 0 \end{bmatrix}, \qquad 10^{-8} \le \mu \le 10^{-6};$$

Map A is a perturbation of the transpose of the companion matrix corresponding to the characteristic polynomial with roots: 10, 1, μ, and $\mu/10$. The μ^2 term in the (2,1) entry of the matrix A was deleted, so that the matrix could be represented with no roundoff error.

The errors $X_i^{computed} - X_i^{exact}$ as a function of the parameter μ, in the computation of the four LCE's at the end of 1000 iterations are shown in Figs. 2a and 2b. The exact LCE's were computed using 100-digit floating-point arithmetic utilizing MAPLE, and the results were then truncated to 16 digits, all other computations were performed using MATLAB within the IEEE floating point standard, i.e., with a machine precision of about 2.2 x 10^{-16}. Figure 2a compares HQRB and GS, and we observe that for small values of μ, $\mu < 10^{-7}$, the performance of GS deteriorates significantly in computing the smallest LCE (the LCE's are ordered as $X_1 > X_2 > X_3 > X_4$). Over the same range of μ the HQRB remains stable. Figure 2b compares HQRB and MGS, here again the largest error occurs in the computation of the smallest LCE using MGS. However the magnitude of this error is negligible when compared to the error obtained from using GS.

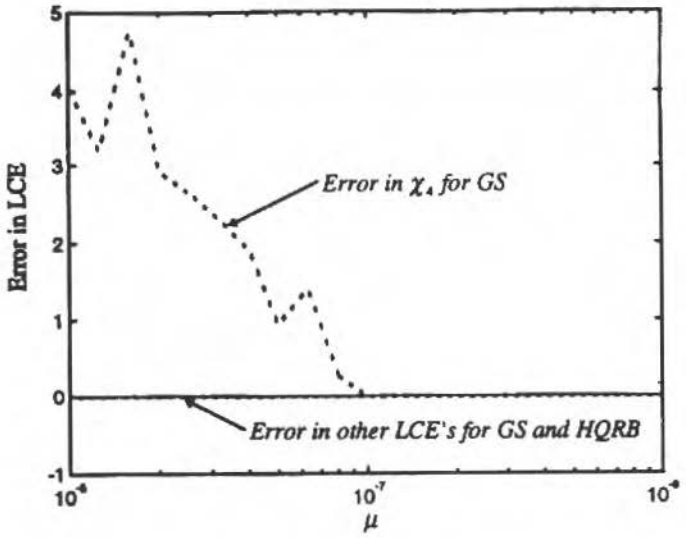


Fig. 2a Error $X_i^{computed} - X_i^{exact}$ versus μ using HQRB and GS for Map A After 1,000 Iterations

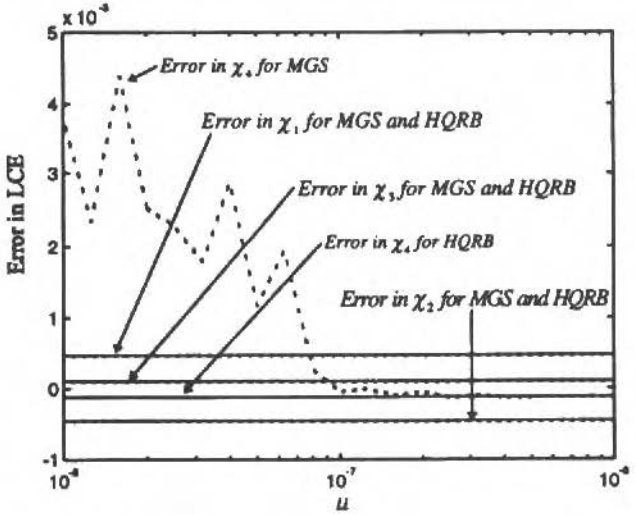


Fig. 2b Error $X_i^{computed} - X_i^{exact}$ **versus** μ **using HQRB and MGS for Map A after 1,000 iterations**

As a measure of accuracy of the methods, we can use the absolute value of the determinant of Q_m, as a measure of the error in orthogonality in the form $c_m = |1 - |Det(Q_m)||$. Departure of c_m from zero would then be an indication of lack of orthogonality. Other measures of error are treated in von Bremen et al. (1997). Figure 3 shows c_m for map A after 1,000 iterations. GS has the largest error in orthogonality, presenting an abrupt change around $\mu = 10^{-7}$. In Figure 2a we can also see that the error in the smallest LCE starts to grow significantly for $\mu < 10^{-7}$. This may be explained with the fact that as μ gets close to 10^{-7} we reach the point where $fl(1+\mu^2) = 1$, and the accumulation of roundoff errors becomes large.

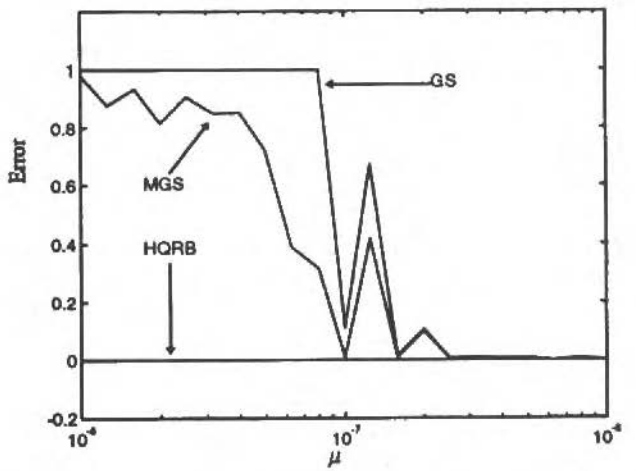


Fig. 3 - Error c_m versus μ **for GS, MGS and HQRB, for Map A after 1,000 iterations**

Conclusions

In this paper we have provided a computationally efficient and robust algorithm for computing the LCE's of a dynamical system. We base our approach on the recognition that (1) approximations to the LCE's can be obtained in a direct manner from the diagonal elements of the R matrix in the QR factorization of the tangent map, and (2) that such a factorization can be efficiently done through a modification in the use of the Householder QR algorithm. The approach proposed here for determining LCE's is shown to be computationally superior to the GS, MGS, and far superior to standard HQR methods. In addition, the numerical experiments reported here show that the algorithm is more stable with respect to roundoff errors than both the GS and the MGS algorithms.

References

Barna, G. and I. Tsuda, A., 1993, "New Method for Computing Lyapunov Exponents", Phys. Lett. A, Vol. 175, No. 6, pp. 421-427.

Benettin, G., Galgani, L., Giorgilli, A. and Strelcyn, J.M., 1978, "Tous les Nombres Characteristiques de Lyapunov Sont Effectivement Calculables", Comptes Rendues Acad. Sc. Paris, Vol. 286A, pp. 431-433.

Benettin, G., Galgani, L., Giorgilli, A. and Strelcyn, J.M., 1980, "Lyapunov Characteristic Exponents for Smooth Dynamical Systems and for Hamiltonian Systems; a Method for Computing all of Them", Part I & Part II, Meccanica, Vol. 15, pp. 9-30.

Dahlquist, G. and Björck, A., 1997, "Numerical Methods" (to appear).

Dieci, L. and Van Vleck, E., 1995, "Computation of a few Lyapunov Exponents for Continuous and Discrete Dynamical Systems", Appl. Numer. Math., Vol. 17, No. 3, pp. 275-291.

Eckmann, J.P. and D. Ruelle, 1985, "Ergodic Theory of Chaos and Strange Attractors", Rev. Mod. Phys., Vol. 57, pp. 617-656.

Geist, K., Parlitz, U. and Lauterborn, W., 1990, "Comparison of Different Methods for Computing Lyapunov Exponents", Prog. Theor. Phys., Vol. 83, No. 5, pp. 875-893.

Golub, G. and Van Loan, C., 1993, "Matrix Computations", 2nd ed., Johns Hopkins University Press.

Parker, T. and Chua, L., 1989, "Practical Numerical Algorithms for Chaotic Systems", Springer-Verlag.

Shimada, I. and Nagashima, T., 1979, "A Numerical Approach to Ergodic Problem of Dissipative Dynamical Systems", Prog. Theor. Phys., Vol. 61, pp. 1605-1616.

von Bremen, H., Udwadia, F. and Proskurowski, W., 1997, "An Efficient QR Based Method for the Computation of Lyapunov Exponents, Physica D (to appear).

Wilkinson, J.H., 1965, "The Algebraic Eigenvalue Problem", Oxford University Press.

Wolf, A., Swift, J., Swinney, H. and Vastano, J., 1985, "Determining Lyapunov Exponents From a Time Series", Physica D, Vol. 16, pp. 285-317.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 262-279

ISSN 0100-7386
Printed in Brazil

Identification of Unstable Mechanical Systems

Roberto Moura Sales
Anselmo Bittar
Michael Porsch
Laércio Lucchesi
Universidade de São Paulo
Escola Politécnica
Depto. de Engenharia Eletrônica - CP 61548
05424-970 São Paulo, SP Brasil
roberto@lac.usp.br

Abstract

In this paper non-linear, unstable, SISO and MIMO mechanical systems are considered. Among three case studies, the first one consists of a lightly damped flexible beam hinged at one end and magnetically levitated at the other end; the second system consists of a magnetically supported rotor, which acts as a water pump in a water tunnel; the third system consists of a magnetically levitated vehicle prototype. Due to the unstable characteristic, several aspects related to closed loop experimental identification of each system are discussed. Analytical and/or experimental models are obtained for each system.
Keywords: *Mechanical Identification, Unstable Systems, Unknown Parameters.*

Introduction

Mechanical systems that use magnetic bearings are frequently employed in many situations related to rotating machines, robotics and more recently for levitated transport systems (Sinha, 1987). This paper is concerned with the identification of three unstable mechanical systems which employ magnetic bearings. The electromagnet systems adopted are inherently unstable, which introduces additional difficulties for any experiment.

The first system is the simplest one in the sense that it is inherently SISO. It consists of a lightly damped flexible beam hinged at one end and magnetically levitated at the other end. The second system consists of a magnetically supported rotor, which acts as a water pump in a water tunnel for hydrodynamic tests; the supporting system for the rotor avoids thus the introduction of turbulence in the water flow. In this case study, resonance frequencies are not relevant, and although it is a MIMO system, the identification procedure could be carried out as in the previous SISO case; the specific characteristics of the rotor resulted in a 5 input-5 output model, particularly suitable for the design of 5 independent controllers. Finally, the third system, a magnetically levitated vehicle prototype, is considered as an inherently MIMO system; in this case, resonance frequencies are present, and too high order models are obtained if some approximations are not taken into account. This system has 4 inputs and 4 outputs.

The paper is outlined as follows: firstly a brief description of each system is presented. The identification procedures and experimental results are then presented for each system. The last section contains general comments and conclusions.

Description of the Systems

Description of the Beam System

The flexible beam considered in this section constitutes a research set up developed for the investigation of different control design techniques when the resonance frequency of the system varies. A schematic diagram of this system is shown in Fig. 1.

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

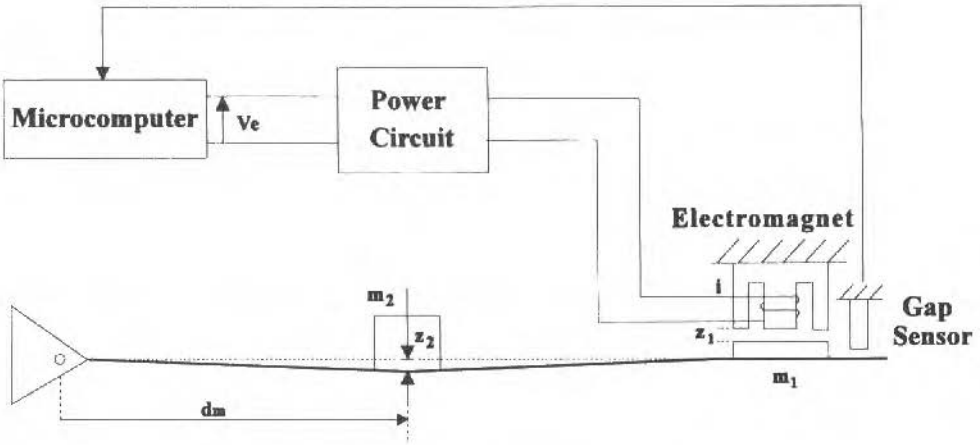


Fig. 1 Schematic Diagram of the Magnetically Levitated Flexible Beam

In this figure, m_1 is a mass of steel through which the magnetic circuit of the electromagnet is closed. The mass m_2 represents a varying load in the middle of the beam, i represents the electrical current, z_1 and z_2 represent the gap of the electromagnet and the deflection of the beam, respectively. The movement of the beam is only in the vertical plane. In order to keep the beam levitated, the gap z_1 of the electromagnet must be controlled. The dimensions of the flexible beam are shown in Table 1.

Table 1 Dimensions of the Flexible Beam

Length	1.28 m
Width	5.08 cm
Thickness	6.35 mm

It is worth to mention some points on the choice of the steady state gap for the electromagnet. Figure 2 shows a schematic diagram of the designed electromagnet.

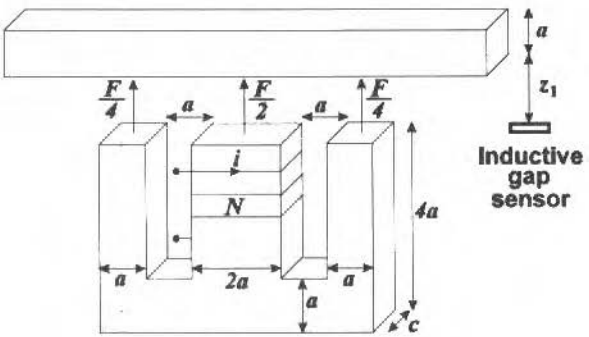


Fig. 2 Electromagnet Scheme

For this type of electromagnet, the attraction force is given by:

$$F(t) = k_f \left(\frac{i(t)}{z_1(t)} \right)^2 \quad (1)$$

where:

i(t): input electrical current,

$z_1(t)$: output air gap,

$k_f = \frac{\mu_0 acN^2}{2}$: constant.

Loosely speaking, the choice of the steady state gap is directly related to the opposite requirements of energy consumption and levitation system robustness. In terms of energy consumption, it is interesting to set the gap as small as possible, in order to get smaller current intensity for levitation. However, for very small gaps, magnetic core saturation may occur, increasing the coil inductance, and leading thus to slow control action; hence, from the point of view of the levitation system robustness, higher gaps are more suitable. In this work, the nominal gap was set as Z_0 = 5mm. This choice is also function of the total mass of the system and electromagnet dimensions. In Table 2 some design data for the electromagnet are presented.

Table 2 Data for the Electromagnet

Nominal air gap	$Z_0 = 5$ mm
Steady state current	$I_0 = 0.7$ A
Dimensions	a = 3 cm c = 6 cm
Number of turns	N=1300
Air permeability	$\mu_0 = 4\rho 10^{-7}$ (H/m)

Description of the Rotor System

This section describes a magnetically supported rotor, which acts as a water pump in a closed circuit water tunnel. Figure 3 shows a schematic diagram of the rotor and its supporting bearings. The purpose of such system is to avoid vibrations through the water flow, which degrades the quality of hydrodynamic tests.

The rotor consists of a 60 cm diameter ring with the propeller in its center; the complete system mass is 75 Kg. The set up is mounted to operate in the vertical plane. The magnetic supporting system consists of 6 electromagnets to levitate and stabilize the ring in the radial plane (electromagnets 7 to 12), and 6 electromagnets to stabilize the rotor in the axial direction (electromagnets 1 to 6).

The electromagnets operate in pairs: 1 - 4, 2 - 5, 3 - 6 in the axial axis and 10 - 11 in the horizontal axis; the only exception is the vertical axis in which the electromagnets 7, 8 and 9 operate with the electromagnet 12, in order to compensate the gravity action. We have thus 5 points to be controlled; a position sensor is placed at each point.

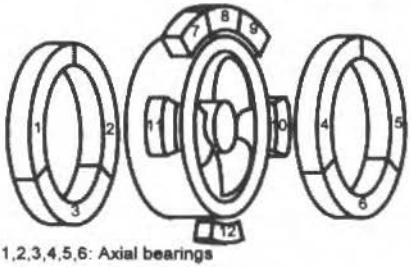


Fig. 3 Schematic Diagram of the Rotor System

Description of the Vehicle System

The constructed system consists of a magnetically levitated vehicle prototype. The levitation is achieved by attraction forces developed by four electromagnets which were positioned in each extremity of the vehicle as shown in Fig. 4. There are also four gap sensors to measure the gaps of these electromagnets.

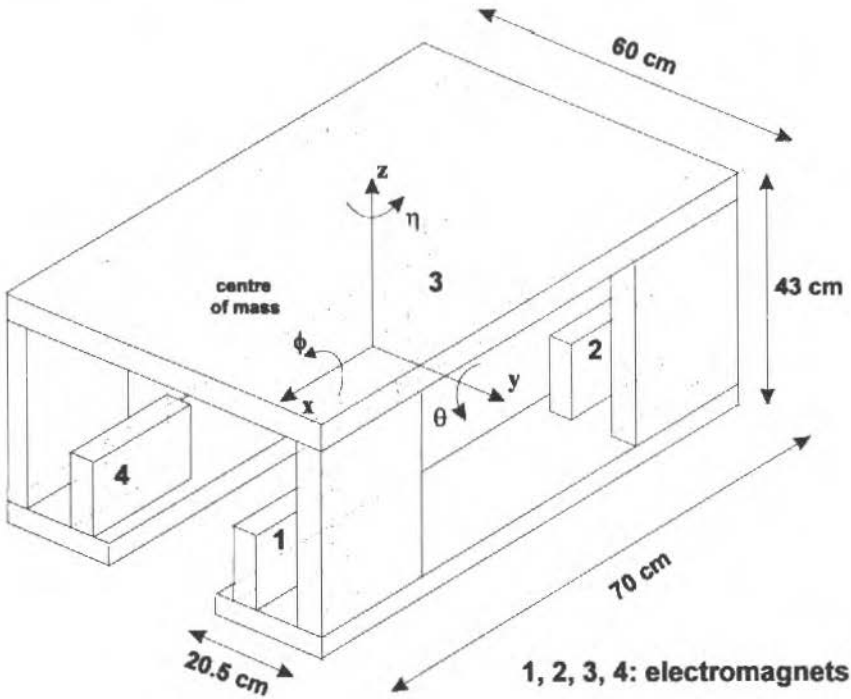


Fig. 4 Schematic Diagram of the Vehicle Prototype

Since the magnetic forces in this case are always attraction forces, an elevated way was also constructed. A vehicle frontal view in levitation state is presented in Fig. 5. The vehicle levitation is achieved through the control of the gaps z_j, j = 1, 2, 3, 4.

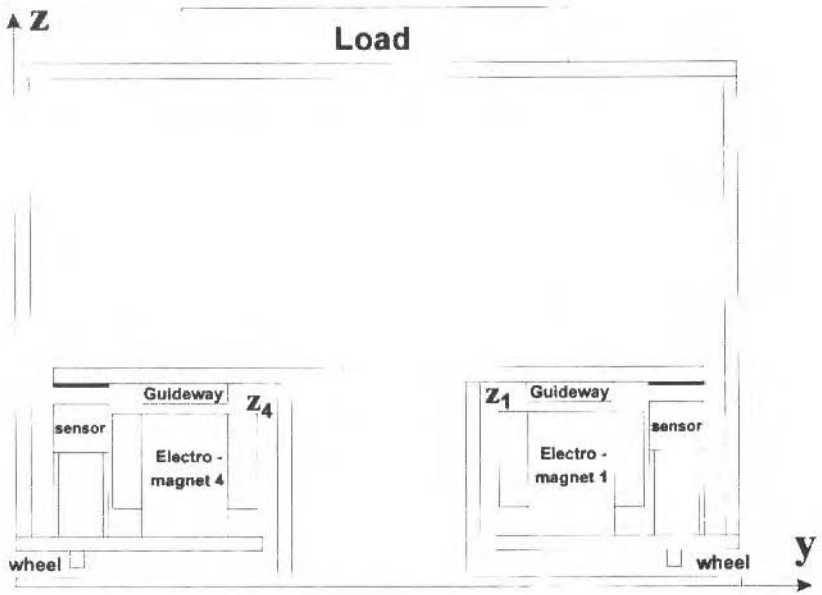


Fig. 5 Vehicle Frontal View

The vehicle has six degrees of freedom which correspond to the three translation motions (x, y, z) and to the three rotation motions (ϕ, θ, η). The vehicle will have independent lift, guide and propulsion systems, but only the levitation system (control of z, ϕ and θ) has been considered in this paper. In order to control the guide system (y e η), it would be necessary to add four lateral electromagnets and to control the respective gaps.

The vehicle was constructed with aluminum plates. The guideways, through which the magnetic flux flows, were constructed with iron bars of small thickness. Laminated guideways would be an alternative choice in order to reduce eddy currents which produce drag and repulsive forces. Drag forces causes power loss and repulsive forces reduce the lift forces (Sinha, 1984). However, iron bars were employed in order to simulate such problems that will be certainly present in any real system. Table 3 presents some prototype data.

Table 3 Prototype Data

Length	70 cm
Width	60 cm
Height	43 cm
Mass	M_V=97.6 kg
Supply Voltage	55 V DC
Levitation power	377 W

Electromagnets for Levitation

The prototype studied in this work has independent electromagnets to lift and to guide the vehicle. This kind of construction provides controllers with simpler structures and improves the reliability of the system (Sinha, 1984). The levitation electromagnets have an "E" shape as in the beam system (Fig. 2), and its design data are also those of Table 2, except the steady state current, I_0, which is equal to 1.71A.

Identification of the Systems

In this section identification procedures and results for each system described before are presented. The common characteristics of the three systems are the fact that they are non-linear and open loop unstable. On the other hand, they present particular features from the point of view of control design. As it will be seen, the flexible beam is an inherently SISO system; in spite of the fact that the rotor is a MIMO system, its particular mechanical characteristics allow the employment of 5 independent controllers, i.e., the plant behaves in fact as a set of SISO decoupled systems; finally, the vehicle prototype will be treated as a MIMO system. Although the unstable characteristic leads to similar identification experiments, the SISO/MIMO behavior imposes specific data treatment.

In the three cases the system to be identified includes the power circuit, the electromagnet, the plant and the sensor. Due to open loop instability, experimental models can be obtained only through closed loop experiments. The first step in the identification procedure was always the adjustment of simple SISO lead-lag controllers in order to stabilize the system. In the case of the beam, such preliminary design consists of the design of a single one input - one output controller; for the rotor and the vehicle 5 and 4 independent SISO controllers, respectively, have to be designed. Some details on the controller designs are presented in Bittar, Sales, Lucchesi and Lima, 1995.

All the experiments are based on the injection of sinusoidal signals in the closed loop. An "HP 3562A Dynamic Signal Analyzer" was employed to inject the sinusoidal signal, and to compute the desired frequency responses as it will be described in the sequel. The noise presence was permanently monitored and for all practical considerations it could be disregarded.

Identification of the Beam System

In order to get some insight into the physics of this system, a simple mathematical model, which can be synthesized in two nonlinear equations relating the involved forces, is presented below - see Fujita, Matsumura and Shimizu, 1990 for more details:

$$m\frac{d^2 z_1}{dt^2} = mg - k_f\left(\frac{i}{z_1}\right)^2 + \alpha(2\cdot z_2 - z_1) + \beta\frac{d}{dt}(2\cdot z_2 - z_1)$$

$$M_b\frac{d^2 z_2}{dt^2} = M_b g - 2\alpha(2\cdot z_2 - z_1) - 2\beta\frac{d}{dt}(2\cdot z_2 - z_1) \qquad (2)$$

where

$$M_b = \frac{m_{beam}}{2} + m_2$$

$$m = \frac{m_{beam}}{2} + m_1$$

k_f - constant proportional to the air permitivity and dependent of the electromagnet dimensions,

α, β - constants that represent forces due to the deflection of the beam,

$k_f\left(\frac{i}{z_1}\right)^2$ - attraction force produced by the electromagnet,

i - electric current,

g - gravity.

The linearization of Eq.(2) produces, in addition to the resonant mode (poles in $\pm 80j$), two real poles (± 70), one stable and one unstable, as shown in Eq.(4). This transfer function includes also one stable pole in -250, which reflects the power circuit dynamic effects and the inductance of the electromagnet (Eq.3). It is assumed that:

$$\frac{I(s)}{V_e(s)} = \frac{k_b}{s+250} \tag{3}$$

where V_e is the input of the power circuit.

This term was estimated experimentally and incorporated in the final transfer function (Eq.4). Since many of the involved constants in Eq.(2) are not easily computed with precision, the obtained transfer function must be taken just as a rough approximation, which will be used for the first controller design.

$$\frac{Z_1(s)}{V_e(s)} = \frac{1690(s+76j)(s-76j)}{(s+250)(s+70)(s-70)(s+80j)(s-80j)} \tag{4}$$

The preliminary stabilizing controller allows to go on the experiments. In this case, the block diagram that represents the stabilized closed loop system for the identification experiments is shown in Fig. 6.

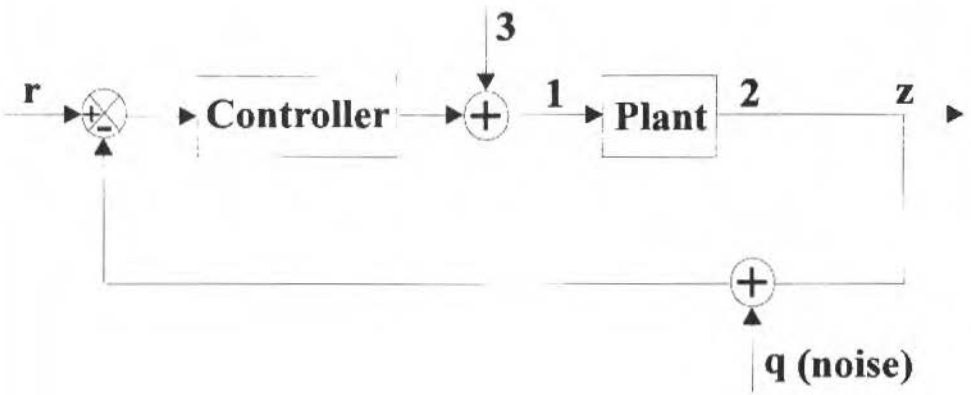


Fig. 6 Closed Loop Block Diagram

Sinusoidal signals of suitable frequencies are injected in point 3. The signals in points 1 and 2 are then measured and their Fourier transforms computed, resulting thus the corresponding frequency response. The injected frequency signal ranges from 10^{-1} Hz to 10^{2} Hz.

It is assumed that m_2 may vary in its position and mass; this assumption leads to specific difficulties since it produces changes in the resonance frequencies. As control specification, it is considered that only the first resonance frequency should be "controlled", in the sense that it will be

the only one included in the controller bandwidth; in addition, the controller should stabilize robustly the plant. These facts have direct implications in the final identified mathematical model.

After measurements are taken, the magnitude and phase plots of $Z_1(jw)/V_e(jw)$ may be obtained. Nine situations are considered accordance to the position d_m and the value of the mass m_2, as shown in Table 4 (the entries of the table represent the model numbers associated to the measurements when the corresponding d_m and m_2 are considered). Thus, as an example, model number 5 corresponds to m_2=1.50kg and d_m=0.71m.

Table 4 Variations of m_2 and d_m

m_2 (kg)	d_m(m)		
	0.565	0.710	0.850
1.25	1	2	3
1.50	4	5	6
1.75	7	8	9

Then, for each data set a curve fit was carried out. In order to fit suitably all data set, a transfer function with 16 zeros and 17 poles was adopted, resulting thus 9 transfer functions. Table 5 presents the poles and zeros of the model number 5.

Table 5 Poles and Zeros of the Model Number 5

Zeros (Hz)	Poles (Hz)
−72.50080	-1.55922
+6.99363	-1.29456
−0.02875 ± 5.38535 · j	+8.47039
+0.02845 ± 34.23650 · j	−0.05321 ± 6.02771 · j
+0.55359 ± 57.12440. j	−0.19195 ± 34.01600 · j
−1.26296 ± 65.78970 · j	−1.41261 ± 65.62790 · j
−0.16784 ± 94.87340 · j	−0.15000 ± 94.83800 · j
+0.13702 ± 99.90610 · j	−0.65930 ± 54.18950 · j
−33.82840 ± 102.46800 · j	+0.14164 ± 99.88560 · j
Gain = -1.09	−25.22210 ± 117.55200 · j

A simple validation procedure for each transfer function can be obtained through the computation of the controller $K_{meas}(jw)$:

$$K_{meas}(jw) = \frac{G_{meas}(jw) - T_{meas}(jw)}{T_{meas}(jw)G_{meas}(jw)} \tag{5}$$

where $G_{meas}(jw)$ represents the measured open loop frequency response from point 1 to point 2, and $T_{meas}(jw)$ represents the measured closed loop frequency response. The designed controller K(jw), which is obviously known, should be equal to $K_{meas}(jw)$ computed from Eq.(5). Figure 7 shows this comparison for the model number 5. The differences are due to difficulties to measure precisely the resonance frequencies (5Hz, 30Hz and 55 Hz, approximately).

The model number 5 (m_2=1.5kg and d=0.710m), for which the values of "d_m" and "m_2" correspond to central values for the admissible variations, was adopted as the nominal one. According to the control specifications, the controller should be robust for variations as in Table 4. Clearly, the order of the fitted nominal transfer function is too high (16 zeros and 17 poles) for control purposes. With this fact in mind, a model reduction algorithm was applied to this nominal transfer function.

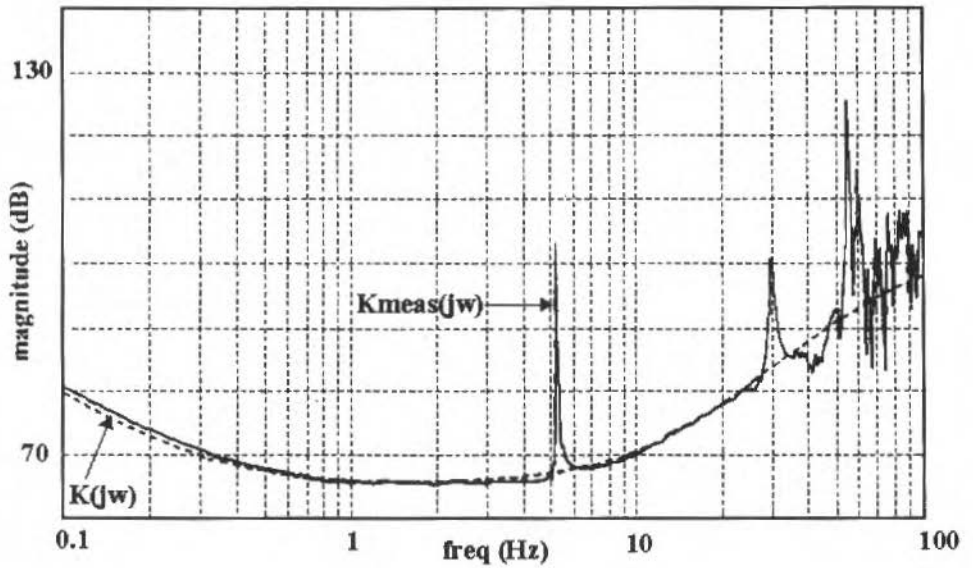


Fig. 7 Comparison Between $K_{meas}(jw)$ and K(jw).

Firstly, it can be observed that if it is desired that the model represents 3 resonance frequencies, then the least necessary degree for the denominator of the transfer function is 9, six of them representing the resonances. Similarly, if it is desired to model 2 resonance frequencies the least degree is 7, and to model 1 resonance frequency the least degree is 5. A mathematical model of order 3 is not able to represent any resonance frequency. Although this model of order 3 represents just the rigid mode of the beam and the electromagnet dynamic, it was chosen for control design. In order to justify this result, let's focus our attention on the modeling uncertainties, by computing the quantities:

$$|\Delta_A(jw)| = |G(jw) - G_N(jw)|$$

and

$$|\Delta_M(jw)| = \left|\frac{G(jw) - G_N(jw)}{G_N(jw)}\right|$$

where

$\Delta_A(jw)$ and $\Delta_M(jw)$ represent additive and multiplicative uncertainties, respectively;

G_N (jw) represents the nominal model after order reduction to 3 poles, (Fig.8), and 5 poles, (Fig.9);

G(jw) represents each one of the models (1, 2, 3, 4, 6, 7, 8, 9 in Table 4) of order 17.

Thus, from Table 4 the respective uncertainties are obtained:

- 8 uncertainty curves of the type $|\Delta_A(jw)|$ with respect to G_N(jw) of order 3;
- 8 uncertainty curves of the type $|\Delta_A(jw)|$ with respect to G_N(jw) of order 5;
- 8 uncertainty curves of the type $|\Delta_M(jw)|$ with respect to G_N(jw) of order 3;
- 8 uncertainty curves of the type $|\Delta_M(jw)|$ with respect to G_N(jw) of order 5.

In Figures (8) and (9) some of these plots are shown.

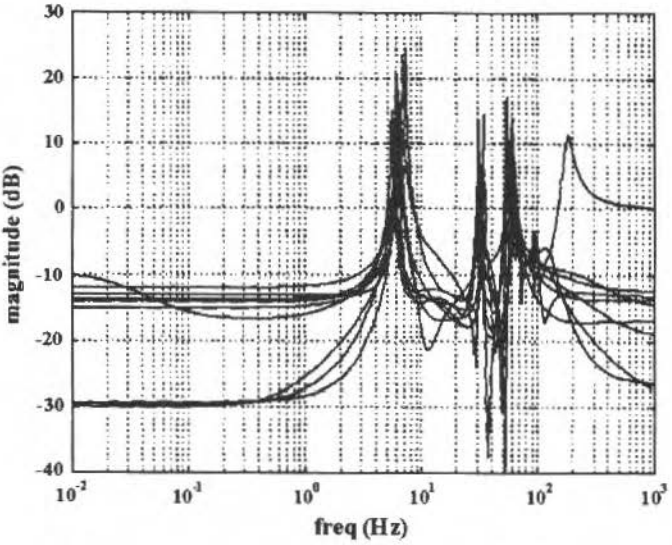


Fig. 8 Multiplicative Uncertainties with Respect to G_N of Order 3.

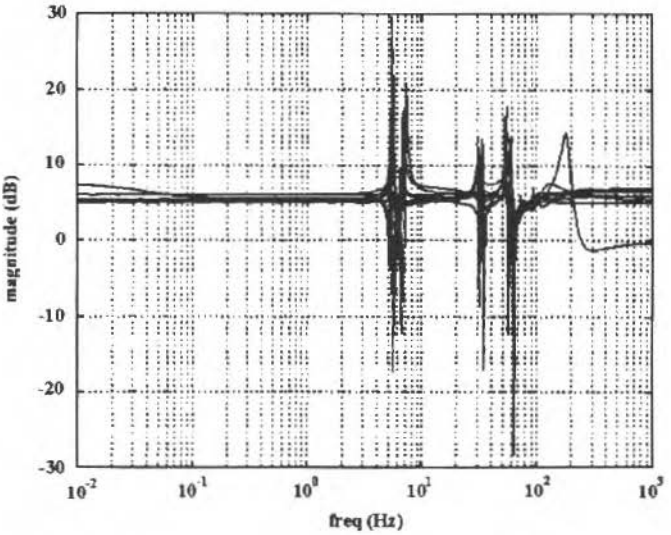


Fig. 9 Multiplicative Uncertainties with Respect to G_N of Order 5.

It is interesting to note that the superior contour for the uncertainty curves are very similar in both figures, above 5 Hz (similar results were obtained from the additive uncertainty plots). This is a consequence of the resonance frequency variation when m_2 and d_m vary. Hence, from the point of view of uncertainty modeling both transfer functions, of order 5 and 3, are equivalent. For simplicity the third order model is recommended for future control design. Equations (6) and (7) present below the transfer functions of order 5 and 3, respectively.

$$G_N(s) = \frac{-7.19(s-408.56)(s+412.90)(s-0.18 \pm 33.83j)}{(s+97.97)(s+81.340)(s-53.22)(s+3.34 \pm 37.87j)} \tag{6}$$

and

$$G_N(s) = \frac{-7.19(s-408.56)(s+412.90)}{(s+107.77)(s+89.47)(s-53.22)} \tag{7}$$

Identification of the Rotor System

For each pair of electromagnets a simple mathematical model can be derived from the schematic diagram of Fig. 10, where a pair of electromagnets applies forces F+ and F- to the mass, M_r, placed between them.

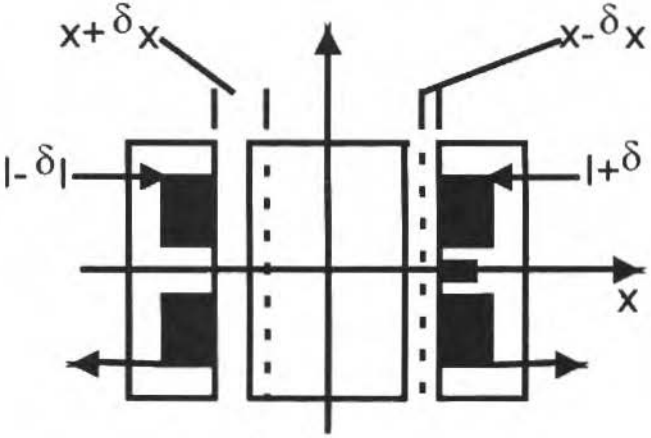


Fig. 10 A Pair of Electromagnets.

In this case,

$$F_+ = K_m \frac{(I+\delta I)^2}{(x-\delta x)^2}$$

$$F_- = K_m \frac{(I-\delta I)^2}{(I+\delta I)^2}$$

and

$$F_r \overset{\Delta}{=} M_r \cdot \delta\ddot{x} = F_+ - F_- = K_m \left[\frac{(I+\delta I)^2}{(x-\delta x)^2} - \frac{(I-\delta I)^2}{(x+\delta x)^2} \right]$$

The state space nonlinear equations result as:

$$x_1 \overset{\Delta}{=} \delta x$$

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = \frac{K_m}{M_r} \left[\frac{(I+\delta I)^2}{(x-\delta x)^2} - \frac{(I-\delta I)^2}{(x+\delta x)^2} \right]$$

and the linearized corresponding equations present two eigenvalues:

$$\lambda_{1,2} = \pm I \sqrt{\frac{K_m}{M_r x^3}}$$

similar to the beam case.

Following this procedure, the complete model was derived by considering each pair of electromagnet independently. In fact, the mechanical characteristics of the rotor suggested that there should have weak couplings between distinct pairs of electromagnets; this was experimentally confirmed as it will be seen in the sequel.

The parameter values for the rotor system are presented in Table 6.

Table 6 Parameter Values for the Rotor System

M_r	75 kg
K_a (D, E and I electromagnets constants)	1.5×10^{-4} Nm/A^2
K_r (X and Y axis electromagnets constants)	1.78×10^{-4} Nm^2/A^2
Gap	0,015 mm

The linear dynamic equations, in the case of the horizontal axis, are:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 2.8 \cdot 10^3 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 4.2 \end{bmatrix} \cdot u$$

$$y = \begin{bmatrix} 0.2 \cdot 10^3 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad (8)$$

where $x_1 = \delta x$, $x_2 = \delta\dot{x}$, $u = \delta I$, and the output matrix represents the position sensor gain (V/m). The equilibrium point is around I = 1 A.

As in the analytical modeling, the experimental measurements were taken independently for each bearing pair. In this case, the experimental procedure follows exactly that one of the beam system.

From Figure 9 it can be noted that the experimental measurements are close to the analytical model (Eq.8), confirming so the weak couplings in the system. Note that, differently from the beam

case, the power circuit and the inductance dynamic effects were neglected in Eq. 8, without introducing significant errors.

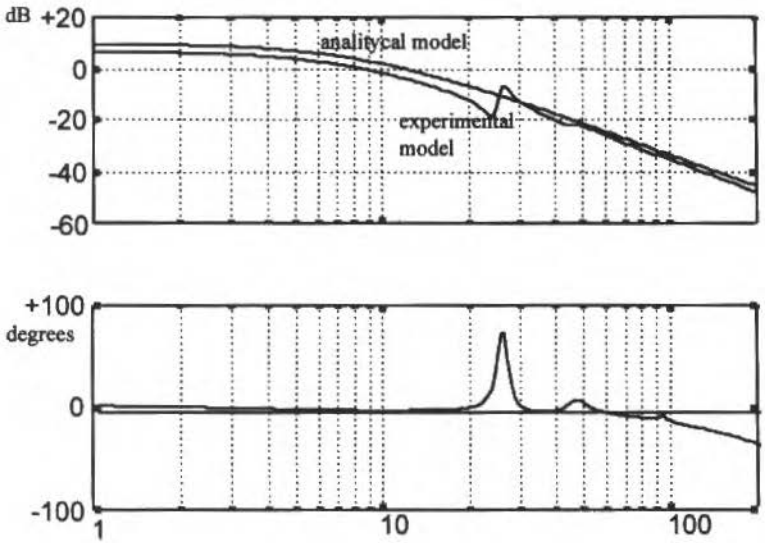


Fig. 11 Horizontal Axis Bode Plots for the Analytical and Experimental Models

Identification of the Vehicle

Although an analytical model could be developed as in Jayawant, Sinha, Wheeler, Whorlow and Willsher, 1976, in this section only the experimental identification procedure and the corresponding results for the vehicle prototype are presented. The point that makes this case different from the above considered systems is the multivariable characteristic. As it will be seen, this fact may lead to very high order models, and some special care must be taken.

As in the rotor case, the first step consists in the design of four independent stabilizing controllers $k_i G_c(s)$ (i=1, 2, 3, 4), implemented as in Fig. 12.

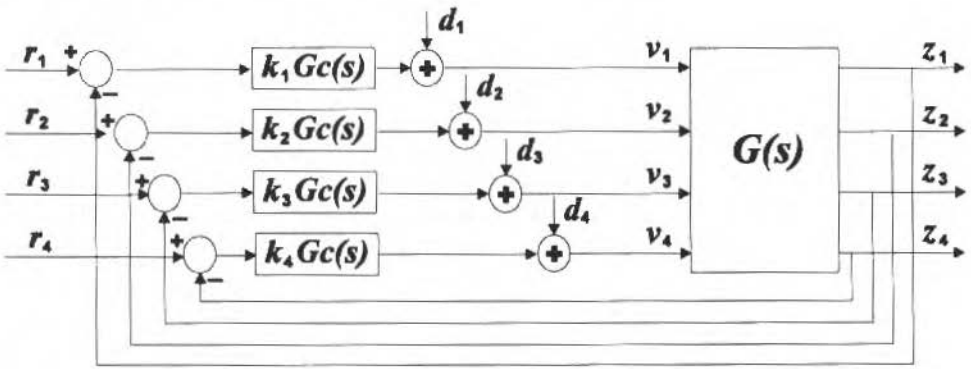


Fig. 12 System Block Diagram

The second step consists in the measurement of the closed loop frequency response. A sinusoidal signal, d_1, is injected as in Fig. 12 and the 4 system outputs, z_1, z_2, z_3 and z_4, are acquired. The

same procedure is repeated separately for the remaining inputs d_2, d_3 and d_4, and thus 16 closed loop frequency responses are obtained.

The same number of inputs and outputs allows some simple operations in order to get an experimental model, i.e.:

$$z_m(jw) \overset{\Delta}{=} (I + G(jw)K(jw))^{-1} G(jw)d(jw) \tag{9}$$

or

$$z_m(jw) \overset{\Delta}{=} G_{mf}(jw)d(jw) \tag{10}$$

and hence

$$G(jw) = G_{mf}(jw)(I - K(jw)G_{mf}(jw))^{-1} \tag{11}$$

for every w such that (I - K(jw)G_{mf}(j w)) is invertible,

In the above equations K(jw) = G_c(jw) diag(k_1 k_2 k_3 k_4), k_i (i = 1, 2, 3, 4) constant, and

$$d(jw) \overset{\Delta}{=} \begin{bmatrix} d_1(jw) \\ d_2(jw) \\ d_3(jw) \\ d_4(jw) \end{bmatrix} \qquad z_m(jw) \overset{\Delta}{=} \begin{bmatrix} z_1(jw) \\ z_2(jw) \\ z_3(jw) \\ z_4(jw) \end{bmatrix}$$

Clearly, the computation of Eq. (11) can be easily performed in the numerical case, i.e., the gain and the phase of G(jw) can be computed for each w. On the other hand, if a curve fit is performed for each experimental frequency response in G_{mf}(jw), then an analytical expression can be obtained for G(jw). However, in this last case too long expressions may occur. This fact can be illustrated for a system with the same number of inputs and outputs, as follows. Let:

ℓ : number of inputs or outputs of the system;
ρ: degree of the denominator of G_c(s);
σ: degree of the common denominator of G_{mf}(s);
$G(jw) \overset{\Delta}{=} G_n(jw)$: open loop frequency response computed numerically through Eq. (11);
$G(s) \overset{\Delta}{=} G_a(s)$: open loop transfer function matrix computed analytically through Eq. (11).

If there is not any zero-pole cancellation, the common denominator degree, ν, of the transfer function matrix G_a(s) is given by:

$$\nu = (\rho + \sigma)\ell$$

In this case, the minimal state space realization (Kailath, 1980) for G_a(s) is given by the McMillan degree, γ, which in the considered problem results as:

$$\gamma = \ell\nu = (\rho + \sigma)\ell^2 \tag{12}$$

Using a controller Gc(s) with ρ=6, Eq.(12) results as:

$$\gamma = (6 + \sigma) \cdot 4^2 = 96 + 16 \cdot \sigma \tag{13}$$

Equation (13) shows that even for small values of σ, the number of open loop states results high, which leads naturally to the application of order reduction algorithms. On the other hand, mechanical models with resonances are frequently ill conditioned, and under these conditions convergence of order reduction algorithms may fail. Fortunately, as a consequence of the geometrical and mechanical symmetries of the vehicle, the 16 identified closed loop frequency responses are very similar; this allowed the approximation shown in the sequel.

In this sense, each one of the 16 closed loop frequency responses were approximated by the same transfer function with degree 4, producing a transfer function matrix $G_{mf}(s)$ with common denominator with degree 4 (σ=4). Figure (14) shows a typical frequency response, which relates the input d_1 to the output z_1, $G_{mf11}(s)$, for the experimental identification and for the adjusted transfer function with degree 4. The rigid body movement of the vehicle in 12rad/s was the main characteristic which was taken into account in this approximation. This curve fit produces, thus, an open loop model, $G_a(s)$, with 160 states (see Eq. 13).

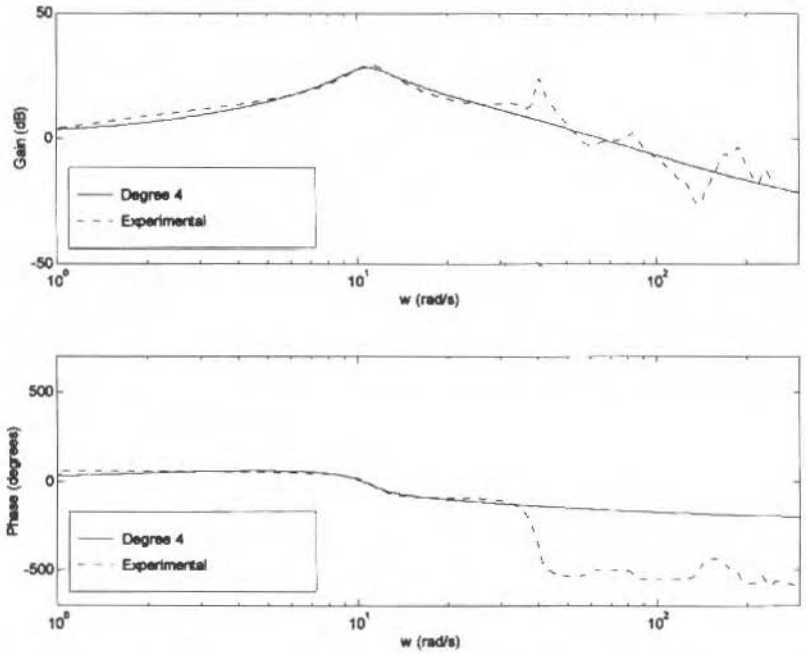


Fig. 13 Frequency Responses of $G_{mf11}(jw)$ for the Experimental Identification and for the approximated transfer function with degree 4.

Using Eq. (11), the computation of the numerical model $G_n(jw)$ and of the analytical model $G_a(s)$ produced the singular values presented in Fig. 14. In order to validate the model $G_a(s)$ with 160 states, step disturbances d_1, d_2, d_3 and d_4 were applied simultaneously to the real and simulated closed loop system. Fig. 15 shows the corresponding time responses.

Although the results in Figs. 14 and 15 may be satisfactory, the identified model $G_a(s)$ with 160 states is considered a high order model for control design, for example. An order reduction algorithm, based on Schur method, was then applied. It is worth to mention that many algorithms were tested, and that due to numerical problems, convergence was achieved only when the state space equations were written using Gilbert's realization (Kailath, 1980). The model of 160 states, $G_a(s)$, was, thus, reduced to a model of 16 states. The singular values for the complete and the reduced order model resulted very similar, as shown in Fig. 16.

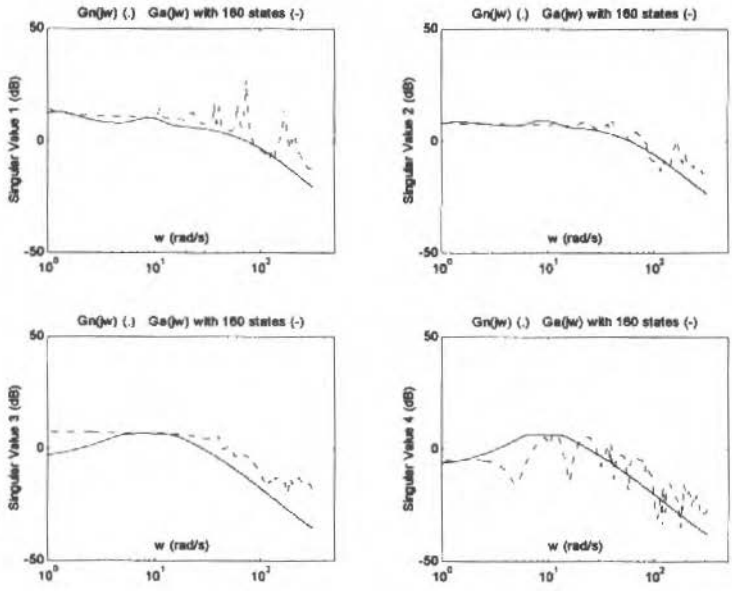


Fig. 14 Singular values of $G_n(jw)$ and $G_a(jw)$.

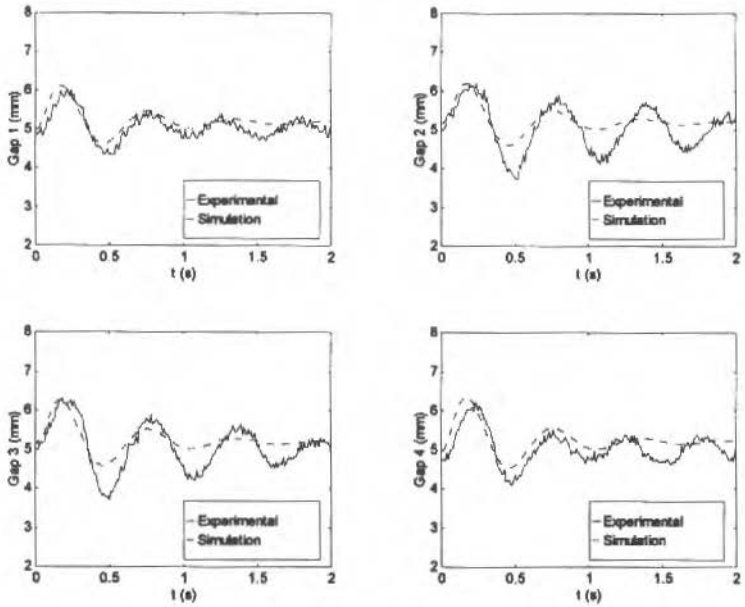


Fig. 15 Closed Loop System Response for a Step Disturbance Applied in Plant Inputs.

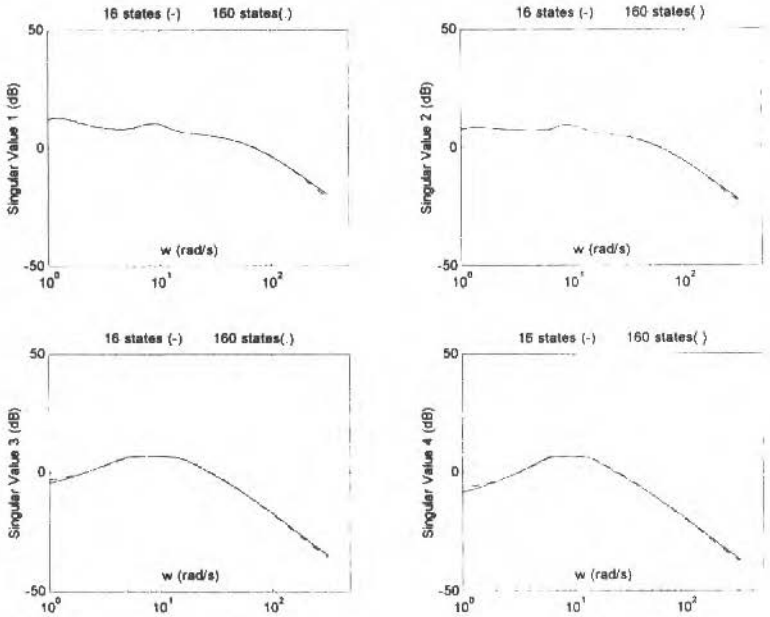


Fig. 16 Singular Values of the Model $G_a(jw)$ With 16 and 160 States.

Equations (14) and (15) present the identified closed loop transfer function matrix, $G_{mf}(s)$, with common denominator of degree 4 and the controller transfer function matrix, K(s), respectively.

$$G_{mf}(s) = -\begin{bmatrix} 1.000 & 0.308 & 0.395 & 1.259 \\ 0.218 & 1.665 & 1.004 & 0.285 \\ 0.390 & 1.191 & 1.221 & 0.549 \\ 0.916 & 0.391 & 0.374 & 1.894 \end{bmatrix} \cdot \frac{0.0256(s+1.6285)(s+10.2255)(s-248.5757)(s+683.0728)}{(s+2.0718\pm 10.5878j)(s+22.0294\pm 2.6509j)} \tag{14}$$

$$K(s) = -\begin{bmatrix} 0.2775 & 0 & 0 & 0 \\ 0 & 0.2667 & 0 & 0 \\ 0 & 0 & 0.2503 & 0 \\ 0 & 0 & 0 & 0.2394 \end{bmatrix} \cdot \frac{(s+0.84)(s+103.27)(s+126.21)(s\pm 7000.00)(s+7834.97)}{s(s+368.42)(s+340.30\pm 178.65j)(s+7196.12\pm 347.33j)} \tag{15}$$

Control Design and Implementation Considerations

In this section some brief comments on control design and implementation are presented. As observed in the "Identification of the Systems" section, the first step in the identification procedure was the adjustment of simple SISO lead-lag controllers to stabilize the plant. For the beam and the vehicle systems, digital controllers to run a sampling frequency of 3 kHz were employed; the rotor controller was analogic.

The final model obtained for the flexible beam was used for H_2, H_∞ and H_2/H_∞ controller design purposes. The main results are presented in (Bittar, Sales, Lucchesi and Lima, 1995). The final model

model obtained for the rotor was used for an analogic H_∞ controller design. The main results are presented in (Porsch, 1996). In the case of the vehicle prototype, the final MIMO system has been the subject of research of H_2 and H_∞ digital controller designs. Robustness and performance are the main characteristics to be investigated.

Concluding Remarks

In this paper identification case studies for non-linear unstable SISO and MIMO mechanical systems were considered. The unstable characteristic imposes that any identification experiment should be carried out in a closed loop configuration.

In the first case study, a flexible beam, emphasis was given to the modeling uncertainty characterization, and the final model, potentially for control design, consisted of a third order model. The system has one input and one output, and the conclusion followed as a direct consequence of the variation of the resonance frequency of the beam. An interesting point to be noted is that, although the resonance frequency is not modeled in the suggested model, such model is suitable for control design even if it is desired that the controller actuates in this frequency range. It seems that some other approach, e.g., μ-synthesis, should be used in order to take into account the effect of the varying resonance.

In the second case study, a magnetically supported rotor, a simple mathematical model was analytically derived. The simplicity of the model relies on the fact that couplings among bearings were neglected. The analytical model was compared with experimental frequency responses and very good results were obtained, confirming so the low coupling assumed.

In the third case study, a magnetically levitated vehicle prototype, an experimental model was obtained. Special emphasis was dedicated to the characterization of the multivariable behavior of the plant. The first experimental model resulted a high order one, and order reduction algorithms had to be employed. In order to get convergence of these algorithms, Gilbert's state space realization developed a decisive role due to ill numerical conditioning of the model. The final 16 states model was validated through its singular values and time responses.

Acknowledgments

This research was supported by FAPESP under grants 91/2344-8 and 95/6968-7.

References

Bittar, A., Sales, R. M., Lucchesi, L., and Lima, A. C., 1995, "Some Design Difficulties and Control Results for a Magnetically Levitated Beam", Proceedings of the third European Control Conference, vol. 4, pp. 3498-3503.

Fujita, M., Matsumura, F., and Shimizu, M., 1990, "H_∞ Robust control design for a magnetic suspension system", 2nd International Symposium on Magnetic Bearing , Tokyo, Japan, July 12-14.

Jayawant, B. V., Sinha, P. K., Wheeler, A. R., Whorlow, R. J., and Willsher, J., 1976, "Development of 1-ton Magnetically Suspended Vehicle Using Controlled D.C. Electromagnets", Proceedings of the IEE, vol. 123, nº. 9, pp. 941-948.

Kailath, T., 1980, "Linear systems", N.J., Prentice-Hall Inc.

Porsch, M., 1996, "Projeto de Sistema de Controle para um Rotor Suportado Magneticamente", Thesis, Departamento de Engenharia Eletrônica, Escola Politécnica da USP, in Portuguese.

Sinha, P.K., 1984, "Design of a Magnetically Levitated Vehicle", IEEE Transactions on Magnetics, vol. MAG-20, nº. 5, pp. 1672-1674.

Sinha, P. K., 1987, "Electromagnetic Suspension: Dynamics and Control", London, Peter Peregrinus Ltd.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 280-290

ISSN 0100-7386
Printed in Brazil

Bounds on Reachable Sets of Dynamical Systems with Uncertain Matrices

Felix Chernousko
Russian Academy of Sciences
117526 Moscow, Russia
chern@ipm.msk.su

Abstract

Linear dynamical systems are considered which are described by ordinary differential equations. We assume that the matrix of the system is uncertain or subject to disturbances, and only the bounds on each element of the matrix are known. We obtain outer ellipsoidal estimates on reachable sets of the system and deduce equations describing the evolution of the approximating ellipsoids. An example is presented. The paper extends the approach of previous papers (Chernousko, 1980, 1982, 1994) where the case of additive disturbances was considered to the more complicated case of multiplicative disturbances. The obtained results make it possible to evaluate disturbances caused by uncertain or perturbed parameters of dynamical systems (e.g., stiffness, damping factors, feedback coefficients, electrical, parameters, etc.).
Keywords: *Dynamical System, Reachable Set, Uncertain Parameters, State Estimation.*

Introduction

Dynamical systems with unknown, uncertain or perturbed parameters arise in many applications. It is important to obtain bounds on reachable sets of such systems, in other words, bounds on all possible motions of these systems. In this paper we consider linear systems of ordinary differential equations with uncertain matrices whose elements are bounded both from below and above. Such systems serve as models for various mechanical, electrical, and other systems whose parameters are not known precisely or can vary in an uncertain way (for example, stiffness and damping factors in mechanical systems, electrical resistance, capacity, and inductivity, feedback coefficients, etc.). We shall obtain outer ellipsoidal estimates on reachable sets and use the method of ellipsoids (see Schweppe, 1973, Kurzhanski, 1977, Chernousko, 1980, 1982, 1994). Ellipsoidal estimates have a number of advantages such as rather simple and explicit form of approximation, smooth boundaries, invariance with respect to linear transformations, etc.

Nomenclature

A = matrix of linear transformation
a = vector of the center of an ellipsoid
a_i = its components
b = vector of translation
b_{ij} = bounds of uncertainties
$C(t)$ = matrix of the linear system
C_0, C_1 = known and unknown parts of C
d = auxiliary vector
c_{ij} = elements of C_1
$D(t, s, M)$ = reachable set
E = ellipsoid
$f(t)$ = given function
G = diagonal matrix
H = auxiliary designation
I = unit matrix
L_1, L_1 = Lagrange's functions
M = initial set
m = number of nonzero b_{ij}
n = order of the system
p = root of the algebraic equation
Q, Q_1, Q_2 = matrices of approxiting ellipsoids
Q_{ij} = elements of matrix Q
q = auxiliary scalar quantity
R^n = n-dimensional space
r_i = semiaxes of an ellipsoid
s = initial instant of time
t = time
^{T} = transposition of a matrix
x = vector of state
x_1, x_2 = auxiliary designations
y = auxiliary vector
y_i = its components
y_i^* = bounds on y_i
Δ = small increment of time
ε = radius of the circle
λ, λ_1, λ_2 = Lagrange's multipliers
λ_j = eigenvalues
μ_j = auxiliary numbers
ν = number of indices for which $r_j = 0$
σ_{ij} = signs of c_{ij}
τ = intermediate instant of time
ξ = auxiliary vector
Ω = set of all possible x_2

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

Throughout, we use the following denotation for an n-dimensional ellipsoid

$$E(a,Q)=\left\{ x:\left(Q^{-1}(x-a),(x-a)\right)\le 1\right\} \tag{1}$$

Here a is an-dimensional center of the ellipsoid, Q is a positive definite n x n-matrix, and the brackets (, , .) denote the scalar product of vectors. Note that the ellipsoid (1) degenerates into a point $x = a$ as $Q \to 0$.

Statement of the Problem

Consider a linear system of ordinary differential equations

$$\dot{x}=C(t)x+f(t) \tag{2}$$

Here, x is an n-dimensional vector of state variables, t is time, C is an n x n-matrix, f is a given n-dimensional vector function of time. We assume that the matrix $C(t)$ consists of two parts: the known part $C_0(t)$ and unknown part $C_1(t)$ which represents uncertainties and disturbances, so that

$$C(t)=C_0(t)+C_1(t) \tag{3}$$

The elements $c_{jk}(t)$ of the matrix $C_1(t)$ satisfy the following inequalities

$$\left|c_{jk}(t)\right|\le b_{jk}(t), \qquad j,k=1,\ldots,n \tag{4}$$

where $b_{jk}(t)$ are given nonnegative functions.

Equations (2) - (4) hold for $t \ge s$ where s is the initial instant of time. The initial state $x(s)$ can be unknown, and only the set M containing this state is given. Thus, we have the following initial condition

$$x(s)\in M, \qquad M\subset R^n \tag{5}$$

The set endpoints $x(t)$ of all state trajectories $x(\cdot)$ of the system (2) under the imposed conditions (3) - (5) is called the reachable set of the system and denoted by $D(t, s, M)$. Thus, we have

$$x(t)\in D(t,s,M) \tag{6}$$

for all $t \ge s$ and all for possible solutions of (2). The reachable set has the following evolutionary property

$$D(t,s,M)=D(t,\tau,D(\tau,s,M)) \tag{7}$$

where τ is any instant such that $\tau \in [s,t]$.

We shall look for outer ellipsoid approximation of reachable set. In other words, we are to find the n-dimensional vector $a(t)$ and $n \times n$-matrix $Q(t)$ (as functions of time t) such that the following inclusion

$$D(t,s,M) \subset E(a(t),Q(t)), \quad t \geq s \tag{8}$$

holds for all $t \geq s$.

Of course, this problem has an infinite number of solutions: any ellipsoid containing the obtained solution $E(a(t),Q(t))$ is also a solution of our problem. It is natural to minimize some measure of approximating ellipsoids, e.g., their volume or the sum of squared semiaxes. This approach was developed in papers (Chernousko, 1980, 1982, 1994) dealing mostly with additive disturbances. Here, we do not obtain optimal approximating ellipsoids but use some optimal operations with ellipsoids so that our approximations for multiplicative disturbances can be called suboptimal.

Transformations of Ellipsoids

Suppose we have obtained the approximating ellipsoid $E(a(t),Q(t))$ satisfying the inclusion (7) for some instant $t \geq s$, so that

$$x(t) \in E(a(t),Q(t)) \tag{9}$$

Let us find the approximating ellipsoid for the instant $t+\Delta$ where $\Delta \geq 0$ is a small increment of time.

It follows from Eqs. (2) and (3) that

$$x(t+\Delta) \approx x_1 + x_2$$

$$x_1 = \left[I + \Delta C_0(t)\right]x(t) + \Delta f(t), \quad x_2 = \Delta C_1(t)x(t) \tag{10}$$

where I is a unit $n \times n$-matrix. According to (10), to obtain the ellipsoid $E(a(t+\Delta),Q(t+\Delta))$ containing the vector $x(t+\Delta)$, it is sufficient to perform the following three operations:

1. Obtain an ellipsoid containing the vector x_1;
2. Obtain an ellipsoid containing the vector x_2;
3. Obtain an ellipsoid containing the sum $x_1 + x_2$.

Let us consider these operations separately.

The first operation is reduced to the affine transformation of an ellipsoid. If some n-dimensional vector x satisfies the inclusion $x \in E(a,Q)$, then

$$Ax+b \in E(Aa+b, AQA^T) \tag{11}$$

Here A is an arbitrary nonsingular $n \times n$-matrix, b is an n-vector, and ^{T} denotes the transposed matrix. Applying the relationship (11) to the vector x_1 from (10), we obtain

$$x_1 \in E(a_1,Q_1), \qquad a_1=\left[I+\Delta C_0(t)\right]a(t)+\Delta f(t)$$

$$Q_1=\left[I+\Delta C_0(t)\right]Q(t)\left[I+\Delta C_0^T(t)\right] \tag{12}$$

Consider now the second operation. Denote by Ω the set of all possible vectors x_2 given by (10) where $x(t)$ satisfies the inclusion (9) and the elements of the matrix $C_1(t)$ obey the constraints (4). It can be proven that, in general, the set Ω is not convex, but it is starlike with respect to the origin of coordinates and symmetrical with respect to all coordinate hyperplanes. To find an ellipsoid containing the set Ω, let us first obtain a rectangular parallelepiped containing Ω. Denote

$$x_2=\Delta y, \quad y=C_1(t)x(t), \quad x(t)=a(t)+\xi \tag{13}$$

Using (4), (9), and (13), we obtain the following estimates

$$|y_i| \le \left|\sum_{k=1}^{n} c_{jk}x_k\right| \le \left|\sum_{k=1}^{n} c_{jk}a_k\right| + \left|\sum_{k=1}^{n} c_{jk}\xi_k\right|$$

$$\le \sum_{k=1}^{n} b_{jk}|a_k| + \max_{c_{jk}} \max_{\xi} \left|\left(c^j,\xi\right)\right| \tag{14}$$

Here, c^j is the n-vector with the components c_{jk}, $k = 1, \ldots, n$. We omit here the argument t and denote by subscripts the respective components of vectors. The maxima in (14) should be taken with respect to c_{jk} and ξ satisfying the constraints (4) and $\xi \in E(0, Q)$, respectively. To find the second maximum in (14), we compose Lagrange's function

$$L=(c^j,\xi)+\lambda(Q^{-1}\xi,\xi)$$

and set its gradient with respect to ξ equal to zero

$$c^j+2\lambda Q^{-1}\xi=0 \tag{15}$$

Here, λ is Lagrange's multiplier. From (15) we get

$$\xi=-(2\lambda)^{-1}Qc^j \tag{16}$$

Substituting (16) into the condition $(Q^{-1}\xi, \xi)$, we obtain the equality $(Qc^j, c^j) = 4\lambda^2$ and determine $\lambda = \pm (1/2)(Qc^j, c^j)^{1/2}$. Inserting this expression for λ into (16), we get

$$\xi = \pm (Qc^j, c^j)^{-1/2} Qc^j \tag{17}$$

We substitute ξ from (17) into (c^j, ξ) and find he second maximum in (14)

$$\max_{\xi} \left| \left(c^j, \xi\right) \right| = \left(Qc^j, c^j\right) = \left(\sum_{p,q=1} Q_{pq} c_{jp} c_{jp} \right)^{1/2} \tag{18}$$

Now, according to (14), we are to maximize the expression (18) with respect to c_{jk} subject to the inequalities (4). Since the quadratic form (Qc^j, c^j) is a convex function of c^j, the desired maximum is attained at one of the apices of the parallelepiped defined by the inequalities (4). We have

$$c_{ij} = b_{ij}\sigma_{ij}, \qquad \sigma_{ij} = \pm 1, \qquad i, j = 1, \ldots, n \tag{19}$$

Taking into account the relationships (14), (18), and (19), we obtain

$$\left| y_j \right| \le y_j^* = \sum_{k=1}^{n} b_{jk} \left| a_k \right| + \left(\max_{\sigma} \sum_{p,q=1}^{n} Q_{pq} b_{jp} b_{jq} \sigma_{jp} \sigma_{jq} \right)^{1/2} \tag{20}$$

Here the maximum is taken over all $\sigma_{ij} = \pm 1$; $i, j = 1, \ldots, n$.

Remark 1. To calculate the maximum in (20), we are to search through 2^{m-1} variants corresponding to various signs of σ_{ij}. Here, m is the number of nonzero elements b_{ij}, and the exponent $m-1$ is conditioned by the fact that the simultaneous change of all signs of σ_{ij} does not change the sum in (20) to be maximized; hence, the sign of one of σ_{ij} can be chosen arbitrarily. In the general case, we have $m = n^2$, but, in many applications, disturbances and uncertainties are present only in a limited number of elements of the matrix C in (2), and m is small.

Remark 2. Let $b_{jk} = 0$ for some j and all $k = 1, \ldots, n$. Then, by virtue of (20), we obtain $y_j^* = 0$, and the set Ω lies in the hyperplane $y_j = 0$. This case occurs, if some rows of the matrix C in (2) do not contain uncertainties.

Remark 3. Let only one element of the matrix C in (2) is uncertain, i.e., $b_{ik} = 0$ except $b_{jp} > 0$. Here, the expression (20) is reduced to explicit formulas

$$y_j^* = b_{jp}\left(\left| a_p \right| + Q_{pp}^{1/2} \right), \qquad y_i^* = 0, \qquad i \ne j \tag{21}$$

Let us now find an ellipsoid containing the rectangular parallelepiped (20). Since the axes of the parallelepiped coincide with the coordinate axes, we take the approximating ellipsoid as follows

$$\sum_{j=1}^{n} r_j^{-2} y_j^2 \le 1 \tag{22}$$

The lengths r_j of the semiaxes are chosen so that the ellipsoid (22) contains the parallelepiped (20). We have

$$\sum_{j=1}^{n} r_j^{-2}\left(y_j^*\right)^2 = 1 \tag{23}$$

We also require that the volume of the ellipsoid (22) (which is proportional to the product of all r_j) should be minimal under the condition (23). Let us compose Lagrange's function corresponding to this conditional extremal problem

$$L_1 = \prod_{j=1}^{n} r_j + \lambda_1 \sum_{j=1}^{n} r_j^{-2}\left(y_j^*\right)^2$$

where λ_1 is Lagrange's multiplier. Setting the derivatives of L_1 with respect to r_j equal to zero, we get

$$r_j^2 = \lambda_2\left(y_j^*\right)^2, \quad \lambda_2 = 2\lambda_1\left(\prod_{j=1}^{n} r_j\right)^{-1}, \quad j = 1,\ldots,n$$

Substituting this expression for r_j^2 into (23), we obtain $\lambda_2 = n$. Thus, we have

$$r_j = n^{1/2} y_j^*, \quad j = 1,\ldots,n \tag{24}$$

If $b_{jk} = 0$ for some j and all $k = 1,\ldots,n$, this result can be improved. In this case, according to Remark 2, we have $y_j^* = 0$. Let the number of indices j such that $b_{jk} = 0$ for all $k = 1,\ldots,n$, is equal to ν, $0 \le \nu < n$. Then the set Ω lies in a $(n - \nu)$-dimensional hyperplane, and it is natural to approximate Ω by a $(n - \nu)$-dimensional ellipsoid lying in the same hyperplane and having minimal $(n - \nu)$-dimensional volume. Instead of (24) we obtain

$$r_j = (n-\nu)^{1/2} y_j^*, \quad j = 1,\ldots,n \tag{25}$$

Now we can present the result of the second operation mentioned at the beginning of Section "*Transformation of Ellipsoids*", i.e., the estimate on the vector x_2 from (13), as follows

$$x_2 \in E(0, Q_2), \quad Q_2 = \Delta^2 G, \quad G = diag(r_1^2,\ldots,r_n^2) \tag{26}$$

Here, the symbol diag{ } denotes the diagonal matrix with the diagonal elements r_j^2, $j = 1,\ldots,n$.

Now we are to find an ellipsoid $E(a(t+\Delta), Q(t+\Delta))$ containing the sum $x(t + \Delta) = x_1 + x_2$ from (10). Here, x_1 and x_2 belong the ellipsoids given by (12) and (26), respectively. Hence, the resulting ellipsoid must satisfy the inclusion $E(a(t+\Delta), Q(t+\Delta)) \supset E(a_1, Q_1) + E(a_2, Q_2)$ where $a_2 = 0$. We shall minimize the volume of the resulting ellipsoid. To this end, we use the formulas for the ellipsoid of the minimal volume containing the sum of two given ellipsoids (Chernousko, 1980, 1982, 1994)

$$a(t+\Delta) = a_1 + a_2 \tag{27}$$

$$Q(t+\Delta)=(p^{-1}+1)Q_1+(p+1)Q_2$$

Here, $p > 0$ is a unique positive root of the algebraic equation

$$\sum_{i=1}^{n}\frac{1}{p+\lambda_i}=\frac{n}{p(p+1)} \tag{28}$$

Here, $\lambda_1 \geq 0$, $i = 1, \ldots, n$, are the roots of the characteristic equation

$$det(Q_1-\lambda Q_2)=0 \tag{29}$$

Each root is counted according to its multiplicity.

Differential Equations for the Evolution of Approximating Ellipsoids

Substituting a_1 from (12) and $a_2 = 0$ into the first formula (27), we obtain

$$a(t+\Delta)=\left[1+\Delta C_0(t)\right]a(t)+\Delta f(t)$$

As $\Delta \to 0$, this equation turns into the differential equation for the function $a(t)$

$$\dot{a}=C_0(t)a+f(t) \tag{30}$$

Now we substitute Q_1 from (12) and Q_2 from (26) into the second formula (27). We get

$$\begin{aligned}Q(t+\Delta)=(p^{-1}+1)\{Q(t)+\Delta[C_0(t)Q(t)+Q(t)C_0^T(t)\\ +O(\Delta^2)\}+(p+1)\Delta^2 G\end{aligned} \tag{31}$$

Let as evaluate the root p of equation (28). Inserting Q_1 and Q_2 from the respective formulas (12) and (26) into Eq. (29), we obtain

$$det\left[Q(t)+O(\Delta)-\Delta^2\lambda G\right]=0 \tag{32}$$

The roots λ_j of equation (32) for small Δ can be presented as follows

$$\lambda_j=\Delta^{-2}(\mu_j)^{-1}+\ldots, \qquad j=1,\ldots,n \tag{33}$$

where μ_j are new unknowns. Here and below, dots denote terms of higher orders in Δ. Substituting (33) into (32), we obtain the following equation for μ_j

$$det\left[Q^{-1}(t)-\mu_j I\right]=0 \tag{34}$$

Since the matrix Q is positive definite and the matrix G is nonnegative definite, Eq. (34) has n nonnegative roots μ_j (each root is counted according to its multiplicity). Let us seek the unique root p of Eq. (28) in the form

$$p = \Delta^{-1} q^{-1} + \ldots \tag{35}$$

Substituting expansions (33) and (35) into Eq. (28) and expanding its both sides into series in Δ, we obtain

$$\Delta^2 \sum_{l=1}^{n} \mu_j (1 + \ldots) = n \Delta^2 q^2 (1 + \ldots)$$

Thus, we have

$$q = n^{-1/2} \left(\sum_{l=1}^{n} \mu_j \right)^{1/2} \tag{36}$$

The sum of all roots of the characteristic equation (34) is equal to the trace of the matrix $Q^{-1}G$. Hence, we obtain from (36)

$$q = \left\{ n^{-1} \left[Tr\left(Q^{-1} G \right) \right] \right\}^{1/2} \tag{37}$$

Inserting (35) into Eq. (31) and omitting terms of order Δ^2, we obtain after transformations

$$Q(t + \Delta) = Q(t) + \Delta \left[C_0(t) Q(t) + Q(t) C_0^T(t) \right] + \Delta q Q(t) + \Delta q^{-1} G(t) + \ldots$$

As $\Delta \to 0$, this equation becomes the differential equation for the matrix $Q(t)$

$$\dot{Q} = C_0(t) Q + Q(t) C_0^T(t) + qQ + q^{-1} G \tag{38}$$

The obtained equations (30) and (38) together with the relationships (26) for G, (25) for r_j, (20) for y_j^*, and (37) for q form a system of differential equations for the vector $a(t)$ symmetrical positive definite matrix $Q(t)$. To obtain initial conditions, we should find an ellipsoid $E(a_0, Q_0)$ containing the initial set M from (5) $(E(a_0, Q_0) \supset M)$ and take

$$a(s) = a_0, \qquad Q(s) = Q_0 \tag{39}$$

The obtained results are summarized in the following theorem.

Theorem. The ellipsoid $E(a(t), Q(t))$, whose parameters $a(t)$ and $Q(t)$ satisfy the differential equations (30) and (38) (together with the relationships (26) for G, (25) for r_j, (20) for y_j^*, and (37) for q) under the initial conditions (39), satisfies the inclusions (8) for $t \geq s$ and (9) for all admissible $x(t)$.

Remark 4. The linear system (30) for $a(t)$ can be integrated independently of the nonlinear system (38) for $Q(t)$, whereas the latter system depends on $a(t)$ through the matrix G, see (26), (25), (20).

Remark 5. The obtained approximating ellipsoids have the evolutionary property similar to the property (7) of reachable sets. Namely, we have

$$E(a(t),Q(t)) \supset D(t,\tau,E(a(\tau)),Q(\tau))), \qquad s \leq \tau \leq t$$

This property directly follows from the procedure described above: the approximating ellipsoid at each instant of time is obtained directly from the ellipsoid for the preceding instant.

Remark 6. The nonlinear system (38) for the matrix Q resembles the similar system for the case of additive disturbances (Chernousko, 1980, 1982, 1994). The difference lies in the expression for the matrix G which depends on a and contains the operation of maximization, see (20).

Remark 7. As it was already mentioned in Remarks 1 - 3, the operation of maximization in (20) can be often considerably simplified.

Example

Let us consider the following two-dimensional system ($n = 2$)

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = -x_1 + c(t)x_1, \qquad |c(t)| \leq b \tag{40}$$

where $c(t)$ is an uncertain disturbance and b is a positive constant. If $c(t)$ is a periodic function, then the system (40) describes the parametrically excited linear oscillator. In our case, there is only one nonzero element c_{21} of the matrix c_1, and the formulas (21) hold with $j = 2$, $p = 1$. On the strength of (26), (25), (21), we have $\nu = 1$ and

$$C_0 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad G = \begin{bmatrix} 0 & 0 \\ 0 & b^2\left(|a_1| + Q_{11}^{1/2}\right)^2 \end{bmatrix} \tag{41}$$

In our case, the system (30) becomes

$$\dot{a}_1 = a_2, \qquad \dot{a}_2 = -a_1 \tag{44}$$

We calculate q from (37) and compose the system (38), taking into account the relationships (41). Thus we obtain

$$\dot{Q}_{11} = 2Q_{12} + qQ_{11}, \qquad \dot{Q}_{12} = Q_{11} - Q_{22} + qQ_{12}$$

$$\dot{Q}_{22} = -2Q_{12} + qQ_{22} + q^{-1}b^2\left(|a_1| + Q_{11}^{1/2}\right)^2 \tag{43}$$

$$q = 2^{-1/2} b Q_{11}^{1/2} \left(|a_1| + Q_{11}^{1/2} \right) \left(Q_{11} Q_{22} - Q_{12}^2 \right)^{-1/2}$$

Let the initial set M at $t = 0$ be a circle with the radius ε and the centre at the origin of coordinates in the $x_1 x_2$ plane. Then, we have

$$a_1(O) = a_1(O) = 0, \quad Q_{11}(O) = Q_{22}(O) = \varepsilon^2, \quad Q_{12}(O) = 0 \tag{44}$$

The system (42) under the initial conditions (44) has the trivial solution $a_1 = a_2 = 0$. Then the system (43) becomes

$$\dot{Q}_{11} = 2Q_{12} + bQ_{12}^2 H^{-1}$$

$$\dot{Q}_{12} = Q_{22} - Q_{11} + bQ_{11} Q_{12} H^{-1} \tag{45}$$

$$\dot{Q}_{22} = -2Q_{12} + bQ_{11} Q_{12} H^{-1} + bH$$

$$H = \left[2 \left(Q_{11} Q_{22} - Q_{12}^2 \right) \right]^{1/2}$$

Note that the right-hand sides of Eqs. (45) are homogeneous functions Q_{ij}, and the equations are invariant with respect to the transformation $Q_{ij} \rightarrow \lambda Q_{ij}$ with the parameter λ. Hence, without loss of generality, we can set $\varepsilon = 1$ in (44).

Results of numerical solution of the initial value problem (45), (44) with $b = 0.8$ are presented in Fig. 1. Knowing the functions $a = 0$ and $Q_{ij}(t)$, we can estimate all possible solutions of the system (40) under all admissible disturbances $c(t)$ on the strength of (9).

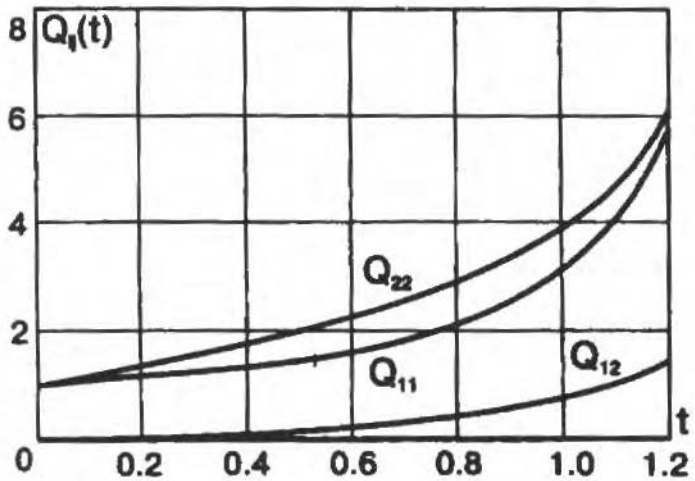


Fig. 1 Time History of Q_{ij}

Conclusion

The estimates presented in the paper make it possible to evaluate the influence of parametric uncertainties and disturbances in linear dynamical systems.

References

Chernousko, F.L., 1980, "Guaranteed Estimates of Undetermined Quantities by Means of Ellipsoids", Soviet Mathematics Doklady, Vol. 21, No. 2, pp. 396-399.

Chernousko, F.L., 1982, "Ellipsoidal Bounds for Sets of Attainability and Uncertainty in Control Problems", Optimal Control Applications and Methods, Vol. 3, No. 2, pp. 187-202.

Chernousko, F.L., 1994, "State Estimation for Dynamic Systems", CRC Press, Boca Raton, Florida, USA.

Kurzhanski, A.B., 1977, "Control and Observation in the Conditions of Uncertainty", Nauka, Moscow (in Russian).

Schweppe, F.C., 1973, "Uncertain Dynamic System", Prentice Hall, Englewood Cliffs, USA.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 291-303

ISSN 0100-7386
Printed in Brazil

The Time-Consistency Problem in Nonlinear Dynamics

Leon Aganesovich Petrosjan
St-Petersburg State University
198904 St-Petersburg, Russia
lapetr@robot.apmath.spb.su

Abstract

The nonlinear dynamic optimization problem (NDOP) $\Gamma(x0,[t0,T])$ on the time interval $[t_0,T]$ from the initial state x0 is considered. If the NDOP is multicriterial or game-theoretic, there exists a rich variety of optimality principles each one consisting from a set of optimal decisions. Denote one such optimality principle in $\Gamma(x_0,[t_0,T])$ by $M(x_0,[t_0,T])$. For a given optimal decision $\overline{m}\ (x_0,[t_0,T]) \in M\ (x_0,[t_0,T])$ let $\overline{x}\ (t)$ be the corresponding optimal trajectory. Consider NDOP subproblems along $\overline{x}\ (t)\Gamma(\overline{x}\ (t),[t,\ T]),\ t\in,[t_0,T]$. Let $M(\overline{x}\ (t),\ [t,\ T])$ be the corresponding optimality principle. Denote by $\overline{m}\ (\overline{x}\ (t),[t,T])$ the trace of optimal decision $\overline{m}\ (x_0,[t_0,T])$ in subproblem $\Gamma(\overline{x}\ (t),[t,\ T])$. The optimality principle is called time-consistent if for every $m\ (x_0,[t_0,T]) \in M\ (x_0,[t_0,T])$ the trace $\overline{m}\ (\overline{x}\ (t),\ [t,\ T]) \in M\ (\overline{x}\ (t),[t,T])$ Most of the classical optimality principles in NDOP $\Gamma(x0,[t0,T])$ (as Shown in (Petrosjan, 1993; Petrosjan and Zenkevich, 1996)) are time inconsistent. The "agreeable solution" (Kaitala and Pohjola, 1988) as an optimality principle in NDOP is considered its time-consistency investigated and the regularization procedure which guarantees the time-consistency of the regularized optimality principle proposed.

***Keywords**: Agreeable Solution, Nash Equilibria, Pareto Optimality, Time-Consistency.*

Introduction

Consider a two-person non-zero-sum differential game $\Gamma(x_0,[t_0,T])$, where the duration $T - t_0$ of the game is fixed but in our case may take also infinite value.

The motion equations are

$$\dot{x} = f(x;u_1,u_2),\ x \in R^n,\ u_i \in U_i \subset R^l,\ i=1,2,t \in [t_0,T]$$

$$x(t_0) = x_0 \tag{1}$$

The payoff of player i is given by:

$$K_i\left(x_0[t_0,T],u_1(\cdot),u_2(\cdot)\right) = \int_{t_0}^{T} e^{-\rho(t-t_0)} g_i\left(x(t),u_1(t),u_2(t)\right)dt,\quad g_i \geq 0,\quad i=1,2 \tag{2}$$

where ζ is the discount rate. We assume that the state trajectory is uniquely defined on $t \in [t_0,T]$ by the initial conditions $x(t_0) = x_0$ and by the strategy pair $u(\cdot) = (u_1(\cdot),u_2(\cdot))$.

If the players agree to cooperate, then they play the open-loop strategy pair $u^*(t) = (u_1^*(t),u_2^*(t))$ which generates a Pareto-optimal (PO) payoff vector

$$\lambda_i\left(x_0[t_0,T]\right) = K_i\left(x_0[t_0,T],u_1^*(t),u_2^*(t)\right) = \int_{t_0}^{T} e^{-\rho(t-t_0)} g_i\left(x^*(t),u_1^*(t),u_2^*(t)\right)dt,\quad i=1,2 \tag{3}$$

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

Denote by $x^*(t)$ the corresponding state trajectory generated as the unique solution of the differential equation:

$$\dot{x}(t)=f(x(t),u_1^*(t),u_2^*(t)), \qquad x^*(t_0)=x_0$$

If the player do not cooperate, then many options are available for the computation of the disagreement point (DP). Indeed, one can assume that the players implement open-loop maximin strategies, or one of Nash equilibria (Nash, 1950) in feedback or open-loop strategies, if it exists. Denote by $\tilde{u}(\cdot)=(\tilde{u}_1(\cdot),\tilde{u}_2(\cdot))$, the chosen disagreement pair of strategies. Let $\tilde{x}(t)$ be the disagreement state trajectory resulting from implementing the disagreement strategy pair $\tilde{u}(\cdot)=(\tilde{u}_1(\cdot),\tilde{u}_2(\cdot))$. Denote by

$$w_i(x_0[t_0,T])=\int_{t_0}^{T} e^{-\rho(t-t_0)}g_i(\tilde{x}^i(t),\tilde{u}_1^i(t),\tilde{u}_2^i(t))dt, \quad i=1,2 \tag{4}$$

the payoff of player i, $i = 1, 2$ obtained for the chosen disagreement strategies.

We restrict our analysis only to those Pareto payoffs that satisfy (global) individual rationality condition, called in the sequel admissible Pareto payoffs, that is

$$\lambda_i(x_0[t_0,T]) \geq w_i(x_0[t_0,T]) \quad i=1,2$$

Consider the family of subgames $\Gamma(x^*(t),[t_0,T])$ along the cooperative Pareto-optimal trajectory $x^*(t)$, $t \in [t_0,T]$. As in $\Gamma(x_0,(t),[t_0,T])$ we introduce the cooperative and disagreement solutions in subgames $\Gamma(x^*(t),[t_0,T])$. From Bellman's optimality principle it follows that the pair of open-loop controls $u(\tau) = (u_1(\tau),u_2(\tau))$, $\tau \geq t$, which coincide with $u^*(\tau)=(u_1^*(\tau),u_2^*(\tau))$ on time interval $[t,T]$, forms a cooperative solution in $\Gamma(x^*(t),[t_0,T])$ (Bellmann, 1957). Corresponding payoffs are given by

$$\lambda_i(x^*(t),[t_0,T])=\int_{t}^{T} e^{-\rho(\tau-t)}g_i(x^*(\tau),u_1^*(\tau),u_2^*(\tau))d\tau, \quad i=1,2 \tag{5}$$

Denote by $w_i(x^*(t),[t,T])$ the noncooperative payoff of player i, $i = 1,2$ in the subgame $\Gamma(x^*(t),[t,T])$. These payoffs are computed with the initial condition $x(t) = x^*(t)$ and by inserting the appropriate pair of strategies (maximin, a feedback or an open-loop Nash) in (4) starting from t.

If the players cooperate from the starting date until the end of the game, then each player will get his (total) Pareto-optimal payoff given by (3). Recall that we have restricted admissible Pareto payoffs to those satisfying (global) individual rationality. This restriction does not guarantee that individual rationality property will be satisfied at any intermediate date. Indeed, it may happen that at a certain t, $t \in [t_0,T]$, one player may find it advantageous to switch to his noncooperative strategy.

The Time Consistency Problem

In this section we construct an allocation mechanism of the Pareto payoffs that insure time consistency or individual rationality at any t, $t \in [t_0,T]$ of the cooperative solution.

Definition 1. The function $\beta_i(t)$ is called a Payoff Distribution Procedure (PDP) if it satisfies

$$\int_{t_0}^{T} e^{-\rho(t-t_0)}\beta_i(t)dt = \lambda_i\left(x_0,[t_0,T]\right), \quad i=1,2 \tag{6}$$

Definition 2. An admissible Pareto-optimal solution is called time consistent if there exist a PDP pair $\beta(t) = (\beta_1(t), \beta_2(t))$ such that:

$$\int_{t}^{T} e^{-\rho(\tau-t_0)}\beta_i(\tau)d\tau \geq e^{-\rho(t-t_0)}w_i\left(x^*(t),[t,T]\right), \quad i=1,2 \tag{7}$$

Proposition 1. An admissible Pareto-optimal solution is time consistent if:

(i) The functions $w_i\left(x^*(t),[t,T]\right), i=1,2$ are differentiable;

(ii) There exist a nonnegative function $\eta_i(t)$, $t \in [t_0,T]$, such that the following conditions are satisfied:

$$\lambda_i\left(x_0,[t_0,T]\right) - w_i\left(x_0,[t_0,T]\right) = \int_{t}^{T} e^{-\rho(t-t_0)}\eta_i(t)dt, \quad i=1,2 \tag{8}$$

Then $\beta_i(t)$ from (7) can be computed by formula

$$\beta_i(t) = \rho w_i\left(x^*(t),[t,T]\right) - \frac{d}{dt}w_i\left(x^*(t),[t,T]\right) + \eta_i(t), \quad i=1,2 \tag{9}$$

Proof. Define the function $\phi_i(t)$ by

$$\int_{t}^{T} e^{-\rho(\tau-t_0)}\beta_i(\tau)d\tau = \phi_i(t), \quad t \in [t_0,T], \quad i=1,2 \tag{10}$$

From (6) it is easy to see that $\phi(t_0) = \lambda_i(x_0,[t_0,T])$. Further, from (6)-(7) it is apparent that

$$\lambda_i\left(x_0,[t_0,T]\right) = \phi(t_0) \geq w_i\left(x_0,[t_0,T]\right)$$

Inequality (7) can be written equivalently

$$\phi(t_0) \geq e^{-\rho(t-t_0)}w_i\left(x^*(t),[t,T]\right), \quad i=1,2 \tag{11}$$

Introduce a nonnegative function $\psi_i(t)$ satisfying

$$\phi_i(t) - \psi_i(t) = e^{-\rho(t-t_0)}w_i\left(x^*(t),[t,T]\right), \quad i=1,2 \tag{12}$$

Suppose $\psi_i(t)$ can be constructed such that

$$\psi_i(t)=\int_t^T e^{-\rho(\tau-\tau_0)}\eta_i(\tau)d\tau, \quad \eta_i(t)\geq 0, \quad i=1,2 \quad \tau\in[t,T] \tag{13}$$

From (12) and (8) we have the following initial condition

$$\psi_i(t_0)=\phi_i(t_0)-w_i(x_0,[t_0,T])=\lambda_i(x_0,[t_0,T])-w_i(x_0,[t_0,T]), \quad i=1,2 \tag{14}$$

Differentiating (12) leads to

$$\beta_i(t)=\rho w_i(x^*(t),[t,T])-\frac{d}{dt}w_i(x^*(t),[t,T])+\eta_i(t), \quad i=1,2 \tag{15}$$

which proves the proposition.

We now construct the function $\eta_i(t)$, that satisfy (8), in the finite and infinite horizon cases.

If T is finite, then we take

$$\eta_i(t)=\frac{\lambda_i(x_0,[t_0,T])-w_i(x_0,[t_0,T]}{T-t_0}e^{\rho(t-t_0)} \tag{16}$$

which obviously satisfies (8).

If the game horizon is infinite, then we take

$$\eta_i(t)=\left\{\lambda_i(x_0,[t_0,T])-w_i(x_0,[t_0,T])\right\} \tag{17}$$

which obviously satisfies (8).

Although the PDP defined guarantees that each player will get exactly his total Pareto-optimal outcome earned for the whole game, it does not insure that $\beta_i(t)$ is nonnegative at each instant of time. In some economic applications it may be undesirable, if for instance borrowing money is not feasible, that $\beta_i(t)$ is negative on some time interval. In the following proposition it is shown that under some conditions we can construct a PDP insuring that $\beta_i(t)$ is nonnegative at any t.

It follows from (9) that the nonnegativity of $\beta_i(t)$ is satisfied if

$$\rho w_i(x^*(t),[t,T])\geq\frac{d}{dt}w_i(x^*(t),[t,T]) \quad t\in[t_0,T], \quad i=1,2$$

Corollary 1. If the discount factor $\rho = 0$, then the condition (9) may be rewritten in a more simple form

$$\beta_i(t)=\eta_i(t)-w_i'(x^*(t),T-t), \quad i=1,2$$

Corollary 2. If the discount factor $\rho = 0$, the condition of the existence of an agreeable solution with nonnegative PDP $\beta_i(t)$ in the case of differentiable $w_i(x^*(t),T-t)$ has the form

$$w_i^t\left(x^*(t), T-t\right) \le 0, \quad t \in [t_0, T] \quad i=1,2$$

Consider now as a cooperative solution such an open-loop strategy pair $\bar{u}(t)=\left(\bar{u}_1(t), \bar{u}_{21}(t)\right)$ and corresponding trajectory $\bar{x}(t)$, which maximize the sum of players payoffs, then we have

$$\sum_{i=1}^{2} K_i\left(x_0, T-t_0; \bar{u}_1(t), \bar{u}_2(t)\right) = \max_{u_1, u_2} \sum_{i=1}^{2} K_i\left(x_0, T-t_0; u_1(t), u_2(t)\right) = v\left(x_0, T-t_0\right) \tag{18}$$

in $\Gamma(x_0, T - t_0)$, and

$$\sum_{i=1}^{2} K_i\left(\bar{x}(t), T-t; \bar{u}_1(t), \bar{u}_2(t)\right) = \max_{u_1, u_2} \sum_{i=1}^{2} K_i\left(\bar{x}(t), T-t; u_1(t), u_2(t)\right) = v\left(\bar{x}(t), T-t\right) \tag{19}$$

in the subgames $\Gamma\left(x(t), T-t\right)$ on the optimal trajectory $\bar{x}(t), t \in [t_0, T]$. Suppose that the payoffs are transferable, then devide the total cooperative payoff $v\left(x_0, T-t_0\right)$, according to the Nash bargaining solution (Haurie, 1976; Nash, 1950) we get in the game $\Gamma\left(x_0, T-t_0\right)$

$$\bar{\lambda}_i\left(x_0, T-t_0\right) = w_i\left(x_0, T-t_0\right) + \frac{v\left(x_0, T-t_0\right) - \sum_{i=1}^{2} w_i\left(x_0, T-t_0\right)}{2}, \quad i=1,2 \tag{20}$$

and in the subgames correspondingly

$$\bar{\lambda}_i\left(\bar{x}(t), T-t\right) = w_i\left(\bar{x}(t), T-t\right) + \frac{v\left(\bar{x}(t), T-t\right) - \sum_{i=1}^{2} w_i\left(\bar{x}(t), T-t\right)}{2}, \quad i=1,2 \tag{21}$$

Definition 3. Suppose $\beta(t) = (\beta_1(t), \beta_2(t))$ is an PDP in the sense that

$$\int_{t_0}^{T} e^{-\rho(t-t_0)} \beta_i(t) dt = \bar{\lambda}_i\left(x_0, T-t_0\right), \quad i=1,2 \tag{22}$$

then PDP $\beta(t)$ is called time consistent if the following condition is satisfied

$$\bar{\lambda}_i\left(x_0, T-t_0\right) = \int_{t_0}^{T} e^{-\rho(\tau-t_0)} \beta_i(\tau) d\tau + e^{-\rho(t-t_0)} \bar{\lambda}_i\left(\bar{x}(t), T-t\right) \tag{23}$$

The following proposition is similar to proposition 1.

Proposition 2. If the functions $v\left(\bar{x}(t), T-t\right), w_i\left(\bar{x}(t), T-t\right), i=1,2$ are differentiable then the time-consistent PDP $\beta(t)$ exists $\left(t \in [t_0, T]\right)$.

Proof. Differentiating (23) with respect to t we get

$$0 = e^{-\rho(t-t_0)}\beta_i(t) + e^{-\rho(t-t_0)}\overline{\lambda}_i'\left(\overline{x}(t), T-t\right) - \rho e^{-\rho(t-t_0)}\overline{\lambda}_i\left(\overline{x}(t), T-t\right) \tag{24}$$

or

$$\beta_i(t) = \rho\overline{\lambda}_i\left(\overline{x}(t), T-t\right) - \overline{\lambda}_i'\left(\overline{x}(t), T-t\right) \tag{25}$$

Using (21) we get from (25)

$$\begin{aligned}\beta_i(t) = \frac{1}{2}\Big\{&\rho\Big[w_i\left(\overline{x}(t), T-t\right) + v\left(\overline{x}(t), T-t\right) - w_{3-i}\left(\overline{x}(t), T-t\right)\Big] - \\ &-\Big[w_i'\left(\overline{x}(t), T-t\right) + v'\left(\overline{x}(t), T-t\right) - w_{3-i}'\left(\overline{x}(t), T-t\right)\Big]\Big\}\end{aligned} \tag{26}$$

Corollary 4. If the discount factor $\zeta = 0$, the formula (28) can be rewritten in a simple form

$$\beta_i(t) = -\frac{1}{2}\left[w_i'\left(\overline{x}(t), T-t\right) + v'\left(\overline{x}(t), T-t\right) - w_{3-i}'\left(\overline{x}(t), T-t\right)\right] \tag{27}$$

Definition 4. The PDP $\beta(t)$ is called strictly time-consistent if in (22), (23) we have an additional condition $\beta(t) \geq 0$.

Using (27) we get the strictly time-consistency conditions for the case $\rho = 0$

$$v'\left(\overline{x}(t), T-t\right) + w_i'\left(\overline{x}(t), T-t\right) \leq w_{3-i}'\left(\overline{x}(t), T-t\right), \quad t \in [t_0, T], \quad i = 1, 2 \tag{28}$$

Regularization

Definition 5. Let the set $A \subset R^m$ be given. The vector $a \in A$ is called weakly Pareto-optimal (WPO) in A, if there does not exist vector $a' \in A$, such that $a' > a$.

We shall denote in this section by $P(x_0, T - t_0)$ the set of all WPO controls in $\Gamma(x_0, T - t_0)$, and by $X(x_0, T - t_0)$ the set of all corresponding trajectories in $\Gamma(x_0, T - t_0)$.

Here we propose a regularization procedure for a weakly Pareto-optimal solution in $\Gamma(x_0, T - t_0)$. Let

$$\lambda(x_0, T-t_0) = \left[\lambda_1(x_0, T-t_0), \lambda_2(x_0, T-t_0)\right]$$

be a weakly Pareto-optimal (WPO) payoff vector corresponding to some WPOS $\overline{u}(t) = \left(\overline{u}_1(t), \overline{u}_2(t)\right)$ in $\Gamma\left(x_0, T-t_0\right)$.

Suppose $\overline{x}(t) \in X\left(x_0, T-t_0\right)$. Denote by $P\left(\overline{x}(t), T-t_0\right)$ the set of all WPO controls in the subgame $\Gamma\left(\overline{x}(t), T-t_0\right)$ and by $X\left(\overline{x}(t), T-t_0\right)$ the corresponding set of all WPO trajectories in $\Gamma\left(\overline{x}(t), T-t_0\right)$.

Let $\overline{\Theta} = \left\{\Theta_1, \Theta_2, \ldots, \Theta_k, \ldots\right\}, \Theta_k \in [t_0, T]$ be an infinite sequence such that $\Theta_k \leq \Theta_{k+1}, k = 1, \ldots, l, \ldots$. Introduce the following controls

$$\hat{u}(t)=\begin{cases}\bar{u}^1(t)\in P(x_0,T-t_0), & t\in[t_0,\Theta_1),\\ \bar{u}^2(t)\in P(\bar{x}^1(\Theta_1),T-\Theta_1), & t\in[\Theta_1,\Theta_2),\\ \ldots\ldots\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\\ \bar{u}^k(t)\in P(\bar{x}^{k-1}(\Theta_{k-1}),T-\Theta_{k-1}), & t\in[\Theta_{k-1},\Theta_k),\\ \ldots\ldots\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\end{cases} \tag{29}$$

if the sequence Θ is infinite, or does not converge, and

$$\hat{u}(t)=\begin{cases}\bar{u}^1(t)\in P(x_0,T-t_0), & t\in[t_0,\Theta_1),\\ \bar{u}^2(t)\in P(\bar{x}^1(\Theta_1),T-\Theta_1), & t\in[\Theta_1,\Theta_2),\\ \ldots\ldots\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\\ \bar{u}^k(t)\in P(\bar{x}^{k-1}(\Theta_{k-1}),T-\Theta_{k-1}), & t\in[\Theta_{k-1},\Theta_k),\\ \ldots\ldots\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\\ \bar{u}(t)\in P(\bar{x}(\Theta),T-\Theta), & t\in[\Theta,T),\end{cases} \tag{30}$$

where $\Theta=\lim\limits_{k\to\infty}\Theta_k$. In (29), (30) $\bar{x}^1(t)\in X(x_0,T-t_0),\bar{x}^2(t)\in X(\bar{x}^1(\Theta_1),T-\Theta_1),\ldots,$ $\bar{x}^k(t)\in X(\bar{x}^{k-1}(\Theta_{k-1}),T-\Theta_{k-1}),\ldots\bar{x}(\Theta)=\lim\limits_{k\to\infty}\bar{x}(\Theta_k),\bar{x}(t)\in X(\bar{x}(\Theta),T-\Theta)$.

Denote by $\hat{x}(t)$ the trajectory corresponding to $\hat{u}(t)=(\hat{u}_1(t),\hat{u}_2(t))$.

For each fixed sequence $\overline{\Theta}$ define the set of all possible control pairs $\hat{u}(t)=(\hat{u}_1(t),\hat{u}_2(t))$, $P_{\overline{\Theta}}(x_0,T-t_0)$, of the form (29), (30).

Let $\overline{P}(x_0,T-t_0)=\bigcup\limits_{\overline{\Theta}}P_{\overline{\Theta}}(x_0,T-t_0)$, and $\overline{X}(x_0,T-t_0)$ the set of all corresponding trajectories. Thus for each $\hat{u}(t)\in\overline{P}(x_0,T-t_0)$ there exist a monotone increasing sequence $\overline{\Theta}=\{\Theta_k\}$ such that $\hat{u}(t)$ is WPO in subgame $\Gamma(\bar{x}(\Theta_k),T-\Theta_k)$ for $t\in[\Theta_k,\Theta_{k+1})$ and $\hat{u}(t)$ is WPO in $\Gamma(\bar{x}(\Theta),T-\Theta)$ for $t\in[\Theta,T)$ if Θ is a limit of $\{\Theta_k\}$.

Definition 5. The set $\overline{P}(x_0,T-t_0)$ is called the set of regularized weak Pareto-optimal controls (RWPO). Each $u(t)\in\overline{P}(x_0,T-t_0)$ is called RWPO control. The state trajectory corresponding to $u(t)\in\overline{P}(x_0,T-t_0)$ is called RWPO trajectory. The set of all RWPO trajectories in $\Gamma(x_0,T-t_0)$ is denote by $\overline{X}(x_0,T-t_0)$.

It is obvious that $\overline{P}(x_0,T-t_0)\supset P(x_0,T-t_0)$ and $\overline{X}(x_0,T-t_0)\supset X(x_0,T-t_0)$, consider e.g. the case when the sequence $\overline{\Theta}$ consists from one point T.

Definition 6. A RWPO control $\hat{u}(t)=(\hat{u}_1(t),\hat{u}_2(t))\in\overline{P}(x_0,T-t_0)$ is called an agreeable solution of $\Gamma(x_0,T-t_0)$ if for all $t\in[t_0,T]$

$$\int_t^T e^{-\rho(\tau-t)}g_i(\hat{x}(\tau),\hat{u}_1(\tau),\hat{u}_2(\tau))d\tau\geq w_i(\hat{x}(t),T-t) \tag{31}$$

for $i=1,2$.

Remark. Definition 6 is a special case of Definition 2 if one in Definition 2 instead of a PO trajectory $\bar{x}(t)$ considers a RWPO trajectory $\hat{x}(t)$, and $\beta_i(t)=e^{-\rho(t-t_0)}g_i\left(\hat{x}(t);\hat{u}_1(t),\hat{u}_2(t)\right)$ (which is always nonnegative).

Suppose that the disagreement payoffs $w(x_0,T-t_0)=\left(w_1(x_0,T-t_0),w_2(x_0,T-t_0)\right)$ correspond to a Nash equilibrium in $\Gamma(x_0,T-t_0)$. Denote the Nash equilibrium by $\tilde{u}(\cdot)=\left(\tilde{u}_1(\cdot),\tilde{u}_2(\cdot)\right)$.

Proposition 3. If the disagreement payoffs correspond to some Nash equilibrium $\tilde{u}(\cdot)$ in $\Gamma(x_0,T-t_0)$ and functions $g_i(x,u_1,u_2) \le M$ are uniformly bounded, then there always exists an agreeable solution of $\Gamma(x_0,T-t_0)$.

Proof. Suppose the $\tilde{u}(\cdot)$ generates a pair of WPO payoffs. Then the state trajectory $\tilde{x}(t)$ corresponding to $\tilde{u}(\cdot)$ is WPO and the open-loop controls $\tilde{u}_i\left(\tilde{x}(t)\right)=\tilde{u}_i(t)$, $i=1,2$ will belong to $P(x_0,T-t_0)$ and hence to $\overline{P}(x_0,T-t_0)$. But along the Nash equilibrium trajectory we have

$$\int_t^T e^{-\rho(\tau-t)}g_i\left(\tilde{x}(\tau);\tilde{u}_1(\tau),\tilde{u}_2(\tau)\right)d\tau = w_i\left(\tilde{x}(t),T-t\right) \tag{32}$$

for all $i = 1,2$ $t \in [t_0,T]$. This means that the WPO open-loop strategy pair $\tilde{u}(t)=\left(\tilde{u}_1(t),\tilde{u}_2(t)\right)$ is an agreeable solution of $\Gamma(x_0,T-t_0)$.

Suppose now that $\tilde{u}(t)=\left(\tilde{u}_1(t),\tilde{u}_2(t)\right)$ is not WPO. Then there exist WPO control pair $\bar{u}^1(t)=\left(\bar{u}_1^1(t),\bar{u}_2^1(t)\right)$ such that the payoffs when using $\bar{u}^1(t)$ strictly dominate the Nash payoffs, i.e.

$$\int_{t_0}^T e^{-\rho(t-t_0)}g_i\left(\bar{x}^1(t);\bar{u}_1^1(t),\bar{u}_2^1(t)\right)dt > w_i\left(\bar{x}_0,T-t\right),\quad i=1,2 \tag{33}$$

There may exist many WPO controls $\bar{u}^1(t)$ satisfying (33).

Denote by $\bar{\bar{x}}^1(t),\bar{\bar{u}}_1^1(t),\bar{\bar{u}}_2^1(t)$ the WPO trajectory and corresponding controls for which

$$\min_i\left[\int_{t_0}^T e^{-\rho(t-t_0)}g_i\left(\bar{\bar{x}}^1(t);\bar{\bar{u}}_1^1(t),\bar{\bar{u}}_2^1(t)\right)dt - w_i(x_0,T-t_0)\right] =$$

$$= \max_{\bar{u}^1=\left\{\bar{u}_1^1(t),\bar{u}_2^1(t)\right\}\in P(x_0,T-t_0)}\left\{\min_i\left[\int_{t_0}^T e^{-\rho(t-t_0)}gi\left(\bar{x}^1(t);\bar{u}_1^1(t),\bar{u}_2^1(t)\right)dt - w_i(x_0,T-t_0)\right]\right\} > 0\,.$$

If also the following condition is satisfied, $t \in [t_0,T]$

$$\int_t^T e^{-\rho(\tau-t_0)} g_i\left(\bar{\bar{x}}^1(\tau), \bar{\bar{u}}_1^1(\tau), \bar{\bar{u}}_2^1(\tau)\right) d\tau \geq w_i\left(\bar{x}^1(\tau), T-\tau\right) \tag{34}$$

$$i = 1, 2,$$

then $\bar{\bar{u}}^1(t) = \left(\bar{\bar{u}}_1^1(t), \bar{\bar{u}}_2^1(t)\right)$ is an agreeable solution.

Suppose on the contrary that (34) does not hold for all t. Since the function on both sides of (34) are continuous in t, there exists an instant $\Theta_1 \in [t_0, T]$ which is the first time at which

$$\int_{\Theta_1}^T e^{-\rho(\tau-\Theta_1)} g_1\left(\bar{\bar{x}}^1(\tau), \bar{\bar{u}}_1^1(\tau), \bar{\bar{u}}_2^1(\tau)\right) d\tau = w_1\left(\bar{\bar{x}}(\Theta_1), T-\Theta_1\right)$$

or

$$\int_{\Theta_1}^T e^{-\rho(\tau-\Theta_1)} g_2\left(\bar{\bar{x}}^1(\tau), \bar{\bar{u}}_1^1(\tau), \bar{\bar{u}}_2^1(\tau)\right) d\tau = w_2\left(\bar{\bar{x}}(\Theta_1), T-\Theta_1\right) \tag{35}$$

Since WPO solution is time consistent (see [6]) the payoff vector

$$\left\{ \int_{\Theta_1}^T e^{-\rho(\tau-\Theta_1)} g_i\left(\bar{\bar{x}}^1(\tau), \bar{\bar{u}}_1^1(\tau), \bar{\bar{u}}_2^1(\tau)\right) d\tau \right\}, \quad i = 1,2$$

is WPO in subgame $\Gamma\left(\bar{\bar{x}}(\Theta_1), T-t_0\right)$. If in the same time the pair of payoffs in Nash equilibrium in subgame $\Gamma\left(\bar{\bar{x}}(\Theta_1), T-t_0\right)$ $w\left(\bar{\bar{x}}(\Theta_1), T-t_0\right)$ constitute a WPO payoff pair, then by the definition the open-loop control $\hat{u}(t) = \left(\hat{u}_1(t), \hat{u}_2(t)\right)$, defined by

$$\hat{u}_i(t) = \begin{cases} \bar{\bar{u}}_i(t), & t \in [t_0, \Theta_1], \\ \tilde{u}_i\left(\tilde{x}(t)\right) = \tilde{u}_i(t), & t \in [\Theta_1, T], \end{cases}$$

$$i = 1,2,$$

where $\bar{\bar{u}}(t)$ is WPO in $\Gamma(x_0, T-t_0)$ and $\tilde{u}(\cdot)$ is Nash equilibrium in $\Gamma\left(\bar{\bar{x}}(\Theta), T-\Theta\right)$ is a RWPO. For $\hat{u}(t)$ and the corresponding trajectory $\hat{x}(t)$ the following conditions are satisfied, $t \in [t_0, T]$,

$$\int_t^T e^{-\rho(\tau-t_0)} gi\left(\hat{x}(\tau), \hat{u}_1(\tau), \hat{u}_2(\tau)\right) d\tau \geq w_i\left(\hat{x}(t), T-t\right), \quad i = 1,2$$

which means that $\hat{u}(t)$ is agreeable solution.

If $w\left(\bar{\bar{x}}^1(\Theta_1), T-\Theta_1\right)$ is not a WPO payoff pair, then there exist a WPO control in $\Gamma\left(\bar{\bar{x}}^1(\Theta), T-\Theta_1\right), \bar{u}^2(t)=\left(\bar{u}_1^2(t), \bar{u}_2^2(t)\right)$ such that the following inequality holds

$$\int_{\Theta_1}^{T} e^{-\rho(\tau-\Theta_1)} g_i\left(\bar{\bar{x}}^2(t); \bar{\bar{u}}_1^2(\tau), \bar{\bar{u}}_2^2(t)\right) dt \geq w_i\left(\bar{\bar{x}}^1(\Theta_1), T-\Theta_1\right), \quad i=1,2 \tag{36}$$

(here $\bar{x}^2(\Theta_1)=\bar{\bar{x}}^1(\Theta_1)$).

There may exist many WPO controls $\bar{u}^2(t)$ satisfying (36).

Denote by $\bar{\bar{x}}^2(t), \bar{\bar{u}}_1^2(t), \bar{\bar{u}}_2^2(t)$ the WPO trajectory and corresponding controls for which

$$\min_i \left[\int_{\Theta_1}^{T} e^{-\rho(t-t_0)} g_i\left(\bar{\bar{x}}^2(t); \bar{\bar{u}}_1^2(t), \bar{\bar{u}}_2^2(t)\right) dt - w_i\left(\bar{\bar{x}}^1(\Theta_1), T-\Theta_1\right) \right] =$$

$$= \max_{\substack{\bar{u}^2=\left\{\bar{u}_1^2(t), \bar{u}_2^2(t)\right\} \in \\ \in P\left(\bar{\bar{x}}^1(\Theta_1), T-\Theta_1\right)}} \left\{ \min_i \left[\int_{\Theta_1}^{T} e^{-\rho(t-t_0)} g_i\left(\bar{x}^2(t); \bar{u}_1^2(t), \bar{u}_2^2(t)\right) dt - w_i\left(\bar{\bar{x}}^1(\Theta_1), T-\Theta_1\right) \right] \right\} > 0.$$

If also the following condition is satisfied for $t \in [\Theta_1, T]$

$$\int_{t}^{T} e^{-\rho(\tau-t_0)} g_i\left(\bar{\bar{x}}^2(\tau); \bar{\bar{u}}_1^2(\tau), \bar{\bar{u}}_2^2(\tau)\right) d\tau \geq w_i\left(\bar{\bar{x}}^2(t), T-t\right), \quad i=1,2 \tag{37}$$

then $\bar{u}^2(t)$ is an agreeable solution in $\Gamma\left(\bar{x}^1(\Theta_1), T-\Theta_1\right)$ and

$$\tilde{u}_i(t)=\begin{cases} \bar{\bar{u}}^1(t), & t \in [t_0, \Theta_1], \\ \bar{\bar{u}}^2(t), & t \in [\Theta_1, T]. \end{cases}$$

is an agreeable solution in $\Gamma(x_0, T-t_0)$, since $\tilde{u}(t) \in \bar{P}(x_0, T-t_0)$ by construction.

Suppose on the contrary that (37) does not hold for all t. Since the functions on both sides of (37) are continuous there exist an instant $\Theta_2 \in [\Theta_1, T]$ which is the first time at which

$$\int_{\Theta_2}^{T} e^{-\rho(\tau-\Theta_2)} g_1\left(\bar{\bar{x}}^2(\tau); \bar{\bar{u}}_1^2(\tau), \bar{\bar{u}}_2^2(\tau)\right) d\tau = w_i\left(\bar{\bar{x}}^2(\Theta_2), T-\Theta_2\right),$$

or

$$\int_{\Theta_2}^{T} e^{-\rho(\tau-\Theta_2)} g_2\left(\bar{\bar{x}}^2(\tau); \bar{\bar{u}}_1^2(\tau), \bar{\bar{u}}_2^2(\tau)\right) d\tau = w_2\left(\bar{\bar{x}}^2(\Theta_2), T-\Theta_2\right),$$

If now $w\left(\bar{\bar{x}}^2(\Theta_2), T-\Theta_2\right)$ is a WPO payoff pair, then by the definition the open-loop controls $\hat{u}(t)=\left(\hat{u}_1(t), \hat{u}_2(t)\right)$, defined by

$$\hat{u}_i(t)=\begin{cases}\bar{\bar{u}}_i^1(t), & t \in [t_0, \Theta_1], \\ \bar{\bar{u}}_i^2(t), & t \in [\Theta_1, \Theta_2], \\ \tilde{u}_i\left(\tilde{x}(t)\right)=\tilde{u}_i(t) & t \in [\Theta_2, T].\end{cases}$$

where $\bar{\bar{u}}^1(t)$ is WPO in $\Gamma(x_0, T-t_0)$, $\bar{\bar{u}}^2(t)$ is WPO in $\Gamma\left(\bar{x}^1(\Theta_1), T-\Theta_1\right)$, $\tilde{u}(\cdot)$ is Nash equilibrium in $\Gamma\left(\bar{x}(\Theta_2), T-\Theta_2\right)$ is a RWPO and the following conditions are satisfied $t \in [t_0, T]$,

$$\int_t^T e^{-\rho(\tau-t_0)} g_i\left(\hat{x}(\tau); \hat{u}_1(\tau), \hat{u}_2(\tau)\right) d\tau \geq w_i\left(\hat{x}(t), T-t\right), \quad i=1,2 \quad t \in [t_0, T],$$

which means that $\hat{u}(t)$ is agreeable solution.

If $w\left(\bar{\bar{x}}(\Theta_2), T-\Theta_2\right)$ is not a WPO payoff pair then we proceed in the same way as we have done at moment Θ_1. Continuing this process we may have the following possibilities:

$$\int_{\Theta_{k-1}}^T e^{-\rho(\tau-\Theta_{k-1})} g_1\left(\bar{\bar{x}}^{k-1}(\tau); \bar{\bar{u}}_1^{k-1}(\tau), \bar{\bar{u}}_2^{k-1}(\tau)\right) d\tau = w_1\left(\bar{\bar{x}}^{k-1}(\Theta_{k-1}), T-\Theta_{k-1}\right), \tag{38}$$

or

$$\int_{\Theta_{k-1}}^T e^{-\rho(\tau-\Theta_{k-1})} g_2\left(\bar{\bar{x}}^{k-1}(\tau); \bar{\bar{u}}_1^{k-1}(\tau), \bar{\bar{u}}_2^{k-1}(\tau)\right) d\tau = w_2\left(\bar{\bar{x}}^{k-1}(\Theta_{k-1}), T-\Theta_{k-1}\right), \tag{39}$$

the payoff vector in Nash equilibrium in subgame $\Gamma\left(\bar{\bar{x}}^{k-1}(\Theta_{k-1}), T-\Theta_{k-1}\right)\left(w_1\left(\bar{\bar{x}}^{k-1}(\Theta_{k-1}), T-\Theta_{k-1}\right), w_2\left(\bar{\bar{x}}^{k-1}(\Theta_{k-1}), T-\Theta_{k-1}\right)\right)$ is WPO.

2) Continuing our construction we never come to the case 1) and in the same time the sequence $\bar{\Theta}=\{\Theta_k\}$ has no limit points (which means that if $T<\infty$ the sequence Θ will consist only from a finite number of points).

3) Continuing our construction we never come to the case 1) and in the same time the sequence $\bar{\Theta}=\{\Theta_k\}$ has a limit point $\lim_{k\to\infty} \Theta_k = \hat{\Theta}$.

It can be easily seen that in the case 1) the pair of controls $\hat{u}(t)=\left(\hat{u}_1(t), \hat{u}_2(t)\right)$

$$\hat{u}_i(t)=\begin{cases}\bar{\bar{u}}_i^{1}(t), & t\in[t_0,\Theta_1],\\ \dots\dots\dots & \dots\dots\dots\\ \bar{\bar{u}}_i^{k-1}(t), & t\in[\Theta_{k-2},\Theta_{k-1}],\\ \tilde{u}_i\left(\tilde{x}(t)\right)=\tilde{u}_i(t) & t\in[\Theta_{k-1},T]\end{cases}$$
$$i=1,2$$

is a needed agreeable RWPO in $\Gamma(x_0,T-t_0)$. Here $\bar{\bar{u}}^{1}(t)\in\bar{P}(x_0,T-t_0),\dots,\bar{\bar{u}}^{l}(t)\in\bar{P}\left(\bar{\bar{x}}^{l-1},T-\Theta_{l-1}\right),\dots l=1,\dots,k$, and $\tilde{u}(\cdot)$ is Nash equilibrium in $\Gamma\left(\bar{\bar{x}}^{k-1}(\Theta_{k-1}),T-\Theta_{k-1}\right)$.

In the case 2) the control pair

$$\hat{u}_i(t)=\begin{cases}\bar{\bar{u}}_i^{1}(t), & t\in[t_0,\Theta_1],\\ \dots\dots\dots & \dots\dots\dots\\ \bar{\bar{u}}_i^{k}(t), & t\in[\Theta_{k-1},\Theta_k],\\ \dots\dots\dots & \dots\dots\dots\end{cases}\qquad(40)$$
$$i=1,2$$

is a needed agreeable RWPO in $\Gamma(x_0,T-t_0)$. Here $\bar{\bar{u}}^{1}(t)\in P(x_0,T-t_0),\dots,\bar{\bar{u}}^{k}(t)\in P\left(\bar{\bar{x}}^{k-1},(\Theta_{k-1})\right),T-\Theta_{k-1},\dots$

The case 3) is more complicated. It can be shown in this case that $w\left(\hat{x}(\Theta),T-\Theta\right)=\left(w_1\left(\hat{x}(\Theta),T-\Theta\right),w_2\left(\hat{x}(\Theta),T-\Theta\right)\right)$ is always WPO. And thus the control pair

$$\hat{u}_i(t)=\begin{cases}\bar{\bar{u}}_i^{1}(t), & t\in[t_0,\Theta_1],\\ \dots\dots\dots & \dots\dots\dots\\ \bar{\bar{u}}_i^{k}(t), & t\in[\Theta_{k-1},\Theta_k],\\ \dots\dots\dots & \dots\dots\dots\\ \tilde{u}_i\left(\tilde{x}(t)\right)=\tilde{u}_i(t) & t\in[\hat{\Theta},T]\end{cases}$$

is a needed RWPO in $G(x_0,T-t_0)$. Proposition is proved.

References

Bellmann, R., 1957, Dynamic Programming, Princeton University Press, Princeton.

Haurie, A., 1976, "A Note on Nonzero-Sum Differential Games with Bargaing Solution", Journal of Optimization Theory and Applications, Vol. 18,1, pp. 31-39.

Kaitala, V., Pohjola, M., 1988, "Optimal Recovery of a Shared Resource Stock: A Differential Game Model with Efficient Memory Equilibria", Natural Resource Modeling, Vol. 3, 1, pp. 191-199.
Nash, J., 1950, "Equilibrium Points in n-person Games", Proc. Nat. Acad. Sci. U.S.A., 36, pp. 48-49.
Nash, J., 1950, "The Bargaing Problem", Econometrica, 21, pp. 155-162.
Petrosjan, L., 1993, Differential Game of Pursuit, World Scientific, Singapore, London.
Petrosjan, L., Zenkevich, N., 1996, Game Theory, World Scientific, Singapore, N.J.

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 304-312

ISSN 0100-7386
Printed in Brazil

About Some Properties of the Statistical Regularities

Ivanenko Victor Ivanovych
Kyiv Mogyla Academy
Kyiv 252070 Ukraine
vivan@ukma.kiev.ua

Zorich Ivan Volodymyrovych
Viborgska 25 apt. 41
Kyiv 252056 Ukraine

Abstract

In this article the notion of G_1-independence of two sequences has been formulated in the terms of their joint statistical regularity. The example which illustrates the construction of the statistical regularity of the sequence, taking values from the finite set, has been given.
Keywords: *Statistical Regularity, G_1-Independence of Two Sequences, Finitely Additive Measure.*

For describing the independence of two sequences which are not necessarily statistically stable the notion of G_1-independence was proposed in the monograph of Ivanenko and Labkovsky, 1990. This notion uses essentially the loss-function $l(.,.)$, where $l(.,.)$ is some real bounded function.

For convenience we cite here the definition of G_1-independence:

Let X and Y be arbitrary non-empty sets, and $\Theta = X^N$, $U = Y^N$ - sets of all possible sequences of their elements.

Definition. We will say, that the sequences $\{\theta_n\}(\theta_n \in X\ \forall_n \in N)$ and $\{u_n\}(u_n \in Y\ \forall_n \in N)$ are G_1-independence or they are in the relation $G_1 \subset U \times \Theta$ (i.e. $(\{u_n\},\{\theta_n\}) \in G_1$ or $\{\theta_n\} \in G_1(\{u_n\})$) if and only if for any $\varepsilon > 0$ we can show the finite subdivisions

$$X = \sum_{i \in I} X_i \ , \ Y = \sum_{j \in J} Y_i \ , \ I \times J = E_0 + E_1$$

of the sets X, Y, and $I \times J$ onto non-crossing subsets when the following conditions are satisfied:

11. $$\sup_{\substack{x_1, x_2 \in X_i \\ y_1, y_2 \in Y_j}} \left| l(x_1, y_1) - l(x_2, y_2) \right| < \varepsilon \ \forall\ (i,j) \in E_1$$

12. $$\lim_{N \to \infty} \left[\frac{1}{N} \sum_{n=1}^{N} I_{x_i}(\theta_n) I_{y_j}(u_n) - \frac{1}{N^2} \sum_{n_1=1}^{N} I_{X_i}(\theta_{n_1}) \times \sum_{n_2=1}^{N} I_{Y_j}(u_{n_2}) \right] = 0 \quad \forall i \in I,\ j \in J$$

13. $$\limsup_{N \to \infty} \left[\sum_{(i,j) \in E_0} \frac{1}{N} \sum_{n=1}^{N} I_{X_i}(\theta_n) I_{Y_j}(u_n) < \varepsilon \right.$$

Presented at the Ninth Workshop on Dynamics and Control - DYNCO, Rio de Janeiro, RJ - August 11-14, 1996.
Technical Editors: Agenor de Toledo Fleury and Hans Ingo Weber.

The notion of G_1-independence is of great importance in the solving the task of minimization the average losses under condition of indifferent uncertainty.

When we say about the average losses under condition of indifferent uncertainty, we mean the next situation: the uncertainty is begotten by the fact that the value of some parameter θ, taking values in the given set Θ, isn't known. There are also given the set of possible decisions U and the loss function $l(\theta,u)$ which determines the loss when the decision is $u \in U$ and we know the value of parameter $\theta \in \Theta$. We consider the sequence of independent choices $\{u_n\}$, and it is supposed that the appropriate sequence of values of the parameter θ - $\{\theta_n\}$ does not depend on the decisions being made. The task consists in the minimization of average losses upon the whole infinite sequence of choices.

The general mathematical apparatus for describing statistical regularity of, so called, sampling directedness was elaborated in the work of Ivanenko and Labkovsky, 1990 (for more details see Ivanenko and Labkovsky, 1990, p. 50). Any usual sequence can be regarded as a particular form of some sampling directedness. This apparatus was applied to elaborate the general decision theory which includes as a particular case the theory of stochastic (bayesian) decision and the theory of minimax decision. But the connection between the property of G_1-independence of two sequences and their joint statistical regularity wasn't established.

Let's introduce some notions:

Let $PF(X)$ be the family of all finitely additive probability distributions on the set of all subsets of the set X, i.e.

$$PF(X)=\Big\{p\in\big(2^X\to[0,1]\big):\ p(X)=1,\ p(A\cup B)=p(A)+p(B\setminus A)\quad \forall\, A,B\subseteq X\Big\}$$

Let $M(X)$ be the Banach space of all real bounded functions on X, $\tau(X)$ - the $*$-weak topology in $PF(X)$, generated by the system of vicinities

$$U^{\varepsilon}_{f_1,f_2,\dots,f_n}(p)=\big\{p'\in PF(X):\ |pf_i-p'f_i|<\varepsilon,\ \forall i\in 1,n\big\},\ n\in N,\ \varepsilon>0,f_1,f_2,\dots,f_n\in M(X),$$

and $pf=\int_X f(x)p(dx)$ (in a natural sense of the integral over the finitely additive measure; for more details see Dunford and Schwartz, 1958).

Let's designate by $P(X)$ the union of all non-empty closed in the sense of topology $\tau(X)$ subsets of the set $PF(X)$.

With every sequence $\{x_n\}\in X^N$ we will associate the set $P(\{x_n\})$ of the limit (in the sense of $\tau(X)$) points of the sequence $\big\{p_x^{(n)};n\in N\big\}$ where

$$p_x^{(n)}(A)=\frac{1}{n}\sum_{i=1}^{n} I_A(x_i)\quad \forall\ A\in 2^X \tag{1}$$

From the compactness of the topological space $(PF(X),\ \tau(X))$ follows that the set $P(\{x_n\})$ is non-empty, and besides this fact, it is obviously that $P(\{x_n\})$ is closed; hence $P(\{x_n\})\in P(X)$. We will call the set $P(\{x_n\})$ the statistical regularity of the sequence $\{x_n\}$.

Theorem 1. The sequences $\{\theta_n\}$ and $\{u_n\}$ are G_1-independent if and only if their joint statistical regularity $P(\{\theta_n\},\{u_n\})$ (i.e. the statistical regularity of the double sequence $\{(\theta_n, u_n)\}_{n=1}^{\infty}$) has the property: for any $\varepsilon > 0$ we can show the finite subdivisions

$$X = \sum_{i \in I} X_i \ , \ Y = \sum_{j \in J} Y_i \ , \ I \times J = E_0 + E_1$$

of sets X, Y, and $I \times J$ onto non-crossing subsets when the following conditions are satisfied:

C1. $$\sup_{\substack{x_1, x_2 \in X_i \\ y_1, y_2 \in Y_j}} \left| I(x_1, y_1) - I(x_2, y_2) \right| < \varepsilon \ \forall \ (i,j) \in E_1$$

C2. For $\forall p_0 \in P\left(\{\theta_n\},\{u_n\}\right)$ and for $\forall (i,j) \in I \times J$:

$$p_0\left(X_i \times Y_j\right) - p_0\left(X_i \times Y\right) \times p_0\left(X \times Y_j\right) = 0$$

C3. For $\forall p_0 \in P\left(\{\theta_n\},\{u_n\}\right)$: $$\sum_{(i,j) \in E_0} p_0\left(X_i \times Y_j\right) < \varepsilon$$

Proof. Let's prove "if" assertion of the theorem.

Let the sequences $\{\theta_n\}$ and $\{u_n\}$ be G_1-independent. Let's show that in this case the conditions of the theorem are fulfilled.

Let's fix some $\varepsilon > 0$ and the corresponding subdivisions $\{X_i\}_{i \in I}$, $\{Y_j\}_{j \in J}$, E_0 , E_1. Then the condition C1 of the theorem coincides with the condition I1 of the definition of G_1-independence. Let's fix arbitrary subsets X_i and Y_j from the subdivisions of sets X and Y respectively. And let $p_0 \in P(\{\theta_n\},\{u_n\})$.

We designate the sequence $\{C_n(X_i, Y_j)\}$.

$$C_N\left(X_i, Y_j\right) = \frac{1}{N} \sum_{n=1}^{N} I_{x_i}(\theta_n) I_{y_j}(u_n) - \frac{1}{N^2} \sum_{n_1=1}^{N} I_{X_i}\left(\theta_{n_1}\right) \times \sum_{n_2=1}^{N} I_{Y_j}\left(u_{n_2}\right) \qquad (2)$$

According to (I2)

$$\lim_{N \to \infty} C_N\left(X_i, Y_j\right) = 0 \quad \forall (i,j) \in I \times J \qquad (3)$$

By the definition of the statistical regularity p_0 is a limit point of the sequence $\{p_n\}$ in the topological space $(PF(X \times Y), \tau(X \times Y))$.

$$p_N(A \times B) = \frac{1}{N} \sum_{k=1}^{N} I_A(\theta_n) I_B(u_n), \quad \forall A \in 2^X, \forall B \in 2^Y \tag{4}$$

Therefore subdirectedness $g\left(g: T \to PF(X \times Y)\right)$ of the sequence $\{p_n\}$ exists that

$$\lim_{T} g_t = p_0 \tag{5}$$

i.e. $g_t \xrightarrow{T} p_0$ in the topological space $(PF(X \times Y), \tau(X \times Y))$. According to the definition of subdirectedness $\exists h: T \to N$ that $g_t = p_{h(t)}$ and $\forall n_0 \in N \; \exists t_0 \in T: \; \forall t \geq t_0 \; h(t) \geq n_0$.

As follows from (5) for any vicinity $U(p_0)$ of point $p_0 \; \exists \; t_{U(p_0)} \in T$ that for $t \geq t_{U(p_0)}$, $g_t \in U(p_0)$.

Let $I_{CxD}(\omega)$ be the indicator function of the set $C \times D$, i.e.

$$I_{CxD}(\omega) = \begin{cases} 1, \; if \; \omega \in C \times D; \\ 0, \; otherwise. \end{cases} \tag{6}$$

Let's regard the vicinity $U^{\delta}_{I_{x_i \times y_j}, I_{x_i \times y}, I_{x \times y_j}}(p_0)$. From (5) follows that for $\forall \; \delta \geq 0 \; \exists \; t_\delta \in T$ that

$g_t \in U^{\delta}_{I_{x_i \times y_j}, I_{x_i \times y}, I_{x \times y_j}}(p_0)$ when $t \geq t_\delta$,

i.e. (7)

$$\left| g_t(X_i \times Y_j) - p_0(X_i \times Y_j) \right| < \delta, \tag{8}$$

$$\left| g_t(X_i \times Y) - p_0(X_i \times Y) \right| < \delta, \tag{9}$$

$$\left| g_t(X \times Y_j) - p_0(X \times Y_j) \right| < \delta, \tag{10}$$

where $g_t(A \times B) = \int_{X \times Y} I_{A \times B}(\omega) dg_t$.

Let's regard the sequence of vicinities $U^{1/m}_{I_{x_i \times y_j}, I_{x_i \times y}, I_{x \times y_j}}(p_0)_{m=1}^{\infty}$. According to (5), $\forall m \in N \; \exists t_m \in T: \; g_t \in U^{1/m}_{I_{x_i \times y_j}, I_{x_i \times y}, I_{x \times y_j}}(p_0) \; \forall t \geq t_m$.

Let $H^m = \{n \mid n \in N, n = h(t)\ \ t \geq t_m,\ t \in T\}$. We define the subsequence of natural numbers $\{N_k\}$,

$$N_1 = min\{n \mid n \in N \ \ and \ \ n \in H^1\},$$
$$N_2 = min\{n \mid n \in N,\ n \in H^2,\ \ and\ n > n_1\},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$N_k = min\{n \mid n \in N,\ n \in H^k,\ \ and\ n > n_{k-1}\},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

The next assertions are true:

$$\begin{aligned} &\lim_{k\to\infty} p_{N_k}(X_i \times Y_j) = p_0(X_i, Y_j), \\ &\lim_{k\to\infty} p_{N_k}(X_i \times Y) = p_0(X_i, Y), \\ &\lim_{k\to\infty} p_{N_k}(X \times Y_j) = p_0(X, Y_j). \end{aligned} \tag{11}$$

And then

$$\lim_{k\to\infty} p_{N_k}(X_i \times Y_j) - p_{N_k}(X_i \times Y), x\, p_{N_k}(X \times Y_j) = p_0(X_i \times Y_j) - p_0(X_i \times Y), x\, p_0(X \times Y_j)$$

As far as $\{C_{N_k}\}$ is a subsequence of $\{C_N\}$ and because of (3), we have $p_0(X_i \times Y_j) - p_0(X_i \times Y), x\, p_0(X \times Y_j) = 0$. Thus we prove C2.

To prove C3, designate $\{D_N\}$,

$$D_N = \sum_{(i,j)\in E_0} \frac{1}{N} \sum_{n=1}^{N} I_{x_i}(\theta_n)\, I_{y_j}(u_n) \tag{13}$$

Then condition 13 of G_1-independence we can express as

$$\limsup_{N\to\infty} D_N < \varepsilon \tag{14}$$

Just as we do, proving C2, we can use the finiteness of E_0 and pick out the subsequence $\{p_k\}$ of the sequence $\{p_k\}$ that

$$\lim_{k\to\infty} \sum_{(i,j)\in E_0} p_{N_k}(X_i \times Y_j) = \sum_{(i,j)\in E_0} p_0(X_i \times Y_j)$$

But from the other hand $\sum_{(i,j)\in E_0} p_{N_k}(X_i \times Y_j) = D_{N_k}$, and therefore

$$\limsup_{k\to\infty} D_{N_k} = \lim_{k\to\infty} D_{N_k} < \varepsilon \tag{15}$$

So we proved that $\sum_{(i,j)\in E_0} p_0\left(X_i \times Y_j\right) < \varepsilon$.

To prove "only if" assertion of the theorem, assume that the conditions of the theorem are satisfied. We have to show that the sequences $\{\theta_n\}$ and $\{u_n\}$ are G_δ-independent.

Let's fix arbitrary $\varepsilon > 0$ and corresponding subdivisions of X, Y, and $I \times J$.

Condition C1 coincide with I1. Let's show that C2 $\Rightarrow$ I1. We will prove from contrary, Assume that $\exists X_i$ and Y_i that

$$\lim_{N\to\infty} C_N\left(X_i \times Y_j\right) \neq 0 \tag{16}$$

Then $\exists\, \alpha \neq 0$ that for some subsequence of natural numbers $\{N_k\}$

$$\lim_{N\to\infty} C_{N_k} = \alpha \neq 0 \tag{17}$$

The sequence of measures $\left\{p_{N_k}\right\}$ has some limit point p_0 (because of compactness of $(PF(X), \tau(X))$), hence $p_0 \in P(\{\theta_n\},\{u_n\})$. The measure p_0 satisfies the condition C2 for $\forall X_i$ and Y_j:

$$p_0\left(X_i \times Y_j\right) - p_0\left(X_i \times Y\right) \times p_0\left(X \times Y_j\right) = 0$$

Just as in the first part of our proof we can pick out the subsequence $\left\{p_{N_{k_m}}\right\}_{m=1}^{\infty}$ that

$$\lim_{m\to\infty} p_{N_{k_m}}\left(X_i \times Y_j\right) = p_0\left(X_i \times Y_j\right),$$
$$\lim_{m\to\infty} p_{N_{k_m}}\left(X_i \times Y\right) = p_0\left(X_i \times Y\right),$$
$$\lim_{m\to\infty} p_{N_{k_m}}\left(X \times Y_j\right) = p_0\left(X \times Y_j\right).$$

Then

$$\lim_{m\to\infty}\left\{p_{N_{k_m}}\left(X_i \times Y_j\right) - p_{N_{k_m}}\left(X_i \times Y\right) x\, p_{N_{k_m}}\left(X \times Y_j\right)\right\} = p_0\left(X_i \times Y_j\right) - p_0\left(X_i \times Y\right) x\, p_0\left(X \times Y_j\right) = 0$$ (and we remember that $C_{N_{k_m}} = p_{N_{k_m}}\left(X_i \times Y_j\right) - p_{N_{k_m}}\left(X_i \times Y\right) x\, p_{N_{k_m}}\left(X \times Y_j\right)$), but $\lim_{m\to\infty} C_{N_{k_m}} = \alpha \neq 0$ by our assumption.

So we obtain the contradiction which proves that our assumption is not true. Thus from C2 $\Rightarrow$ I2.

We can prove that I3 follows from C3 just as we prove C2 $\Rightarrow$ I2.

The previous theorem makes possible for us to consider the task of making decision under indifferent uncertainty in the light of general decision theory (see Ivanenko and Labkovsky, 1990). It is obviously as well that the tool of statistical regularity is very useful for the studying mass events which are not necessarily statistically stable.

Unfortunately, generally the structure of the statistical regularity is very complicated. But it turns out that one can build the statistical regularity of the sequence which takes values from the finite set (i.e. the set has finite quantity of elements).

Let X be the set which consists of finite quantity of elements, i.e. $X = \{x_1, x_2, ..., x_m\}$. Let $\{\theta_n\}$.

Theorem 2. A finitely-additive measure $p \in PF(X)$ belongs to $P(\{\theta_n\})$ if and only if the sequence of measures $\{p_N\}\left(p_N(A)=\frac{1}{N}\sum_{i=1}^{N} I_A(\theta_i) \text{ for } \forall A \in 2^X \right)$ contains the subsequence $\{p_{N_k}\}$ that $\lim_{k\to\infty} p_{N_k}(A) = p(A) \forall A \in 2^X$.

Proof. To prove the "if" assertion of our theorem we should consider the subsequence of the vicinities $\{V_k(p)\}, V_k(p)=\{p' \in PF(X) \mid |pf - p'f| < \frac{1}{k} \ \forall f \in \{I_A(\omega): A \in 2^X\}\}$. Just as in the first part of the proof of the Theorem 1 it can be shown that exists $p_{N_k} \in V_k(p) \ \forall k \in N$. Therefore $\lim_{k\to\infty} p_{N_k}(A) = p(A)$ *for* $\forall A \in 2^X$.

Let's proof "only if" assertion of the theorem. Suppose that for some $p \in PF(X)$: $\lim_{k\to\infty} p_{N_k}(A) = p(A)$ *for* $\forall A \in 2^X$. It is obvious that any bounded function $f: X \to R$ can be represented in the following form

$$f(\omega) = \lambda_1 I_{x_1}(\omega) + \lambda_2 I_{x_2}(\omega) + ... + \lambda_m I_{x_m}(\omega), \text{ where } \lambda_i = f(x_i)$$

$$I_{x_i}(\omega)=\begin{cases} 1, \text{ if } \omega = x; \\ 0, \text{ otherwise.} \end{cases}$$

$$pf = \int_X f \, dp = \int_X \left(\lambda_1 I_{x_1} + \lambda_2 I_{x_2} + ... + \lambda_m I_{x_m}\right) dp = \sum_{i=1}^{m} \lambda_i \times p(x_i) \qquad (18)$$

Just as in (18) $p_{N_k} f = \sum_{i=1}^{m} \lambda_i \times p_{N_k}(x_i)$.

According to the conditions of the theorem $\lim_{k\to\infty} p_{N_k}(x_i) = p(x_i)\ \forall\, x_i \in X$, therefore $p_{N_k} f \to pf$ when $k \to \infty$, i.e. p is a limit point of the subsequence $\{p_{N_k}\}$. This means that p is a limit point of $\{p_N\}$, i.e. $p \in P(\{\theta_n\})$.

The proved theorem allows us to build the statistical regularity of $\{x_N\}$ when the set X consists of not many elements. In this case we can pick out from $\{p_N\}$ all subsequences that for any of them and for any $A \in 2^X$ the limit $\lim_{k\to\infty} p_{N_k}(A)$ do exists. It follows from the theorem that the set $P(\{\theta_n\})$ coincide with the set of measures which are the limits of these subsequences.

Example

Let's regard the following sequence which takes values from the set $X = \{0,1\}$:

0 1 0 1 0 0 1 1 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 ..... and so on;

the rule of construction of this sequence:

the sequence consists of the series of zeros and ones; series of ones follows series of zeros; the length of series of ones is equal to the length of the previous series of zeros; every next series of zeros has the length that is equal to common length of all previous series of zeros; the first series is a series of zeros which consists from one element.

It is easy to see that the frequency of "0" ($p_n(\{0\})$ - frequency of $\{0\}$) when $n \to \infty$ is in the next limits:

$$\frac{1}{2} \le p_n(\{0\}) \le \frac{2}{3},$$

and for the frequency of "1" we have $p_n(\{1\}) = 1 - p_n(0)$.

The sequence of the frequency of "0" $(\{p_n(0)\})$ has such limit points:

$$p^1_{m,k}(\{0\}) = \frac{2^m + k}{2*2^m + k},\quad p^2_{m,k}(\{0\}) = \frac{2^m}{3*2^m + k},$$

$m = 0,1,2,3, \ldots,\ k \in [0,2^m] \cap Z$;

and for the $\{p_n(\{1\})\}$ we have:

$$p^1_{m,k}(\{1\}) = 1 - p^1_{m,k}(\{0\}) \text{ and } p^2_{m,k}(\{1\}) = 1 - p^2_{m,k}(\{0\})$$

respectively, where Z is a set of integer numbers.

Hence the statistical regularity of our sequence is:

$$P\left(\{x_n\}\right)=\left\{P^1_{m,k}(\cdot)\right\}_{m\in\{0,1,2,3,\ldots\},k\in[0,2^m]\cap Z;}\cup\left\{P^2_{m,k}(\cdot)\right\}_{m\in\{0,1,2,3,\ldots\},k\in[0,2^m]\cap Z;}$$

References

Dunford, N. and Schwartz, J.T., 1958, "Linear Operators. Part 1: General Theory", Interscience Publisher, New York, London.

Ivanenko, V.I. and Labkovsky, V.A., 1990, "The Uncertainty Problem in Decision Making", Naukova Dumka, Kyiv, [In Russian].

Ivanenko, V.I. and Labkovsky, V.A., 1993, "On the Studying of Mass Phenomena", 12th World Congress IFAC, Preprints of Papers, Vol. 9, pp. 451-454.

Kolmogorov, A.N., 1983, "On the Logical Foundation of Probability", Lect. Notes Math., N 1021, pp. 1-5.

Contents continued

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997

ISSN 0100-7386
Printed in Brazil

Abstracts

Guenther, G., and Pieri, E. R., 1997, "Cascade Control of Hydraulic Actuators", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 108-120.

In this paper the problem of hydraulic positioning system is taken into account. It is shown that by modeling the actuator in a convenient form the system can be interpreted as a mechanical subsystem driven by a hydraulic one. This structure is called a cascade control strategy. Cascade controllers based on the linear an nonlinear models are synthesized using Lyapunov direct method. Simulation results illustrated the main characteristics of the cascade control strategy.

Keywords: Hydraulic Control Systems, Hydraulic Actuators, Positioning Systems.

Fleury, A. T., Lopes, J. A., Moscati, N. R., Nigro, F. E. B., and Trielli, M. A., 1997, "Modeling and Simulation Results for a Natural Gas Internal Combustion Engine Coupled to a Hydraulic Dynamometer", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 121-137.

When considering the problem of pollutant emission reduction from internal combustion engines, many technical challenges arise. One of them is the control strategy formulation to be applied to the fuel injection and ignition systems, with the aim at regulating the air-fuel ratio around the stoichiometric value. One of the steps towards this goal is the development of an engine model. This model has to be validated in a testbed, which includes a dynamometer and a gas analyser. IPT Engines Lab is equipped with a hydraulic dynamometer, whose dynamics is slower than the engine one. Modeling and simulation results for these two coupled models antecipate results for testbed operation and helps to understand and control the engine transient, where the dynamometer dynamics is to be excluded from the achieved experimental results.

Keywords: Internal Combustion Engine, Hydraulic Dynamometer, Mathematical Modeling, Engine Testbed Simulation.

Rios Neto, A., 1997, "Stochastic Optimal Linear Parameter Estimation and Neural Nets Training in Systems Modeling", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 138-146.

Supervised training of feedforward neural networks for nonlinear mapping and dynamical systems modeling is addressed.Viewing neural nets training as a stochastic parameter estimation problem, results in Kalman filtering are adapted to develop training algorithms. Many levels of approximation are considered to develop a range of full non parallel to simplified parallel processing versions of algorithms, together with an adaptive approach intended to give to these algorithms the features of good numerical behavior and of distributing the extraction of learning information to all training data.

Keywords: Neural Nets Supervised Training, Neural Nets and Systems Modeling , Stochastic Optimal Estimation Training Algorithms, Kalman Filtering Training Algorithms.

Mook, D. T., and Luton, J. A., 1997, "Numerical Simulations of Interactions Among Aerodynamics, Structural Dynamics, and Control System", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 147-153.

A numerical simulation of the interactions among the structure of an aircraft wing, the flow around it, and the devices that control the deflections of the ailerons is described. In the present simulation, the structure, flowing air, and controls are considered to be the elements of a single dynamic system. All of the governing equations are numerically integrated simultaneously and iteractively. The procedure is illustrated by an example of a very high-aspect-ratio, very flexible wing. Instead of a simple formula for the aerodynamic forces, there is s rather involved computer code. The input to this code, needed to impose the boundary conditions on the flowfield, is the velocity and position of all the points on the wing. As the airspeed increases and the angle of attack decreases in such a way that the lift force remains constant, the uncontrolled wing eventually begins to flutter. When the controls are turned on, the flutter can be suppressed up to approximately twice the critical airspeed.

Keywords: Flow Structure Interaction, Flight Control System.

Espíndola, J. J., and Bavastri, C. A., 1997, "Viscoelastic Neutralizers in Vibrations Abatement: A Nonlinear Optimization Approach", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 154-163.

A general procedure for the optimization of the parameters of dynamic neutralizers is presented. It can be applied to the minimization of the vibration response and sound radiation of linear structures subjected to excitations in a specified frequency range. Modal theory and generalized equivalent quantity concept for the neutralizers, introduced by Espíndola and Silva (1992), are applied to a non-linear optimization scheme. The proposed procedure can be applied to relaxed and time invariant structures. It is not dependent on the structure complexity and the degree of discretization adopted. In such conditions, a significant reduction in computing work is achieved, if compared with the more traditional methods.

Keywords: Vibration Isolation, Viscoelastic Neutralizers, Multidegree of Freedom Vibration.

Bachschmid,N., and Dellupi, R., 1997, " Malfunction Identification in Rotor Systems from Bearing Measurements Using Partial Models of the System", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 164-175.

A method for the identification of different malfunctions which cause only or mainly 1x rev. vibration components is presented. The methodology is based on the model of the shaft alone, therefore avoiding the need of a linearized model of the oil film and of a reliable model of the casings and foundation, and uses the vibration readings in the bearings during coast-down transients. The results show that the method seems to be appropriate to distinguish between different causes such as concentrated unbalances, coupling misalignments and concentrated or distributed bows which could be produced by a partial rub (in a seal e.g.) or by a non uniform heating or cooling transient (in a generator or a steam turbine during load variations), and to determine the location along the rotor, the angular phase and the amount of unbalance or bow, in other words to identify the position and the severity of the malfunction.

Keywords: Identification, Model Based Diagnosis, Rotordynamics.

Corless, M., and Tu, J., 1997, " A Simple State/Uncertainty Estimator for a Class of Uncertain Systems", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 176-191.

We consider here a class of uncertain systems consisting of a nominally linear part and an uncertain/nonlinear/time-varying part which can be regarded as a state-dependent/time-varying "disturbance input". Using only a measured output, we present simple estimators which can asymptotically estimate to any desired accuracy the system state and the disturbance input.

Keywords: Mechanical Uncertain Systems, Nonlinear State Estimation, Nonlinear Filters.

Rill, G., 1997, "Vehicle Modeling for Real Time Applications", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 192-206.

In the automotive industry enhanced control systems are more and more developed by using hardware or software in the loop techniques. For such applications an enhanced nonlinear vehicle model withreal time capacity is necessary. The paper presents a multi-purpose vehicle model where real time application was made possible not by simplifying the model but by using a special model technique, by adopting the generation of the equations of motion to the specific problems in vehicle dynamics, and by using a modified implicit Euler formalism for the numerical solution.

Keywords: Vehicle Dynamics, Real-Time Simulation.

Soize, C., and Le Fur, O., 1997, "Weakly Nonlinear Second-order Dynamical Systems Identification Using a Random Parameters Linear Model", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 207-216.

The objective of this paper is to present an identification procedure which is based on the use of a stochastic linearization method with random coefficients. The model is then defined as a multidimensional linear second-order dynamical system with random coefficients. An optimization procedure is developed to identify the parameters of the probability law of the random coefficients. The identification procedure is described step by step. Finally, an example is presented and shows the interest of the method proposed.

Keywords: System Identification, Stochastic Linearization, Identification Algorithm.

Flashner, H., and Efrati. T., 1997 " Tracking of Mechanical Systems Using Artificial Neural Networks", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 217-227.

A method for tracking control of mechanical systems using artificial neural networks is proposed. The proposed control law consists of a proportional and derivative control action, and a two-layer feedforward neural network used for on-line approximation of the nonlinear part of the system dynamics. Tuning of the neural network's weights is formulated in terms of a constrained optimization problem and solved on-line using a projection method. It is shown that the proposed control law yields closed-loop tracking error that tends asymptotically to zero while the control effort is minimized. The resulting algorithm has a simple structure and requires a very modest computation effort. The problem of tracking control for a two-degree of freedom planar manipulator is used to demonstrate the proposed method.

Keywords: Control, Mechanical Systems, Neural Networks.

Rochinha, F. A., and Sampaio, R., 1997, "A Consistent Approach to Treat the Dynamics of Flexible Systems", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 228-241.

The dynamics of flexible systems, such as robot manipulators, mechanical chains or cables, is becoming increasingly important in engineering. The main question arising from the numerical modeling of large displacements of multibody systems is an appropriate treatment for the large rotations. In the present work an alternative approach is proposed leading to a time-stepping numerical algorithm which achieve stable solutions combined with high precision. In particular, in order to check the performance of the proposed approach, two examples having preserved constants of the motion are presented.

Keywords: Flexible Mechanical Systems, Multibody System Dynamics.

Chen, Y.H., 1997, "Control Design for Uncertain Systems Under Fuzzy Disturbance", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 242-252.

The problem of designing controls for dynamic systems under input disturbance is considered. We consider two possible characterizations of the input disturbance: (1) it is bounded by a crisp number, or (2) it is bounded by a fuzzy number. The control design is purely deterministic. However, the resulting system performance is interpreted differently, depending on the bound information. It may be deterministic or fuzzy (i.e., with a spectrum of possible outcome).

Keywords: Uncertain Systems, Fuzzy Set Theory.

von Bremen, H. F., Udwadia, F. E., and Proskurowski, W., 1997, "Computation of Lyapunov Exponents using Householder Factorization", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 253-261.

An efficient and numerically stable method to determine all the Lyapunov characteristic exponents of a dynamical system is presented. The method is compared with known methods in terms of efficiency, and the accuracy and stability of the methods are tested by numerical experiments.

Keywords: Nonlinear Dynamical Systems, Lyapunov Exponents, Householders Factorization.

Sales, R. M., Bittar, A., Porsch, M., and Lucchesi, L., 1997 "Identification of Unstable Mechanical Systems", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 262-279.

In this paper non-linear, unstable, SISO and MIMO mechanical systems are considered. Among three case studies, the first one consists of a lightly damped flexible beam hinged at one end and magnetically levitated at the other end; the second system consists of a magnetically supported rotor, which acts as a water pump in a water tunnel; the third system consists of a magnetically levitated vehicle prototype. Due to the unstable characteristic, several aspects related to closed loop experimental identification of each system are discussed. Analytical and/or experimental models are obtained for each system.

Keywords: Mechanical System Identification, Unstable Systems, Unknown Parameters.

Chernousko, F., 1997, "Bounds on Reachable Sets of Dynamical Systems With Uncertain Matrices", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 280-290.

Linear dynamical systems are considered which are described by ordinary differential equations. We assume that the matrix of the system is uncertain or subject to disturbances, and only the bounds on each element of the matrix are known. We obtain outer ellipsoidal estimates on reachable sets of the system and deduce equations describing the evolution of the approximating ellipsoids. An example is presented. The paper extends the approach of previous papers (Chernousko, 1980, 1982, 1994) where the case of additive disturbances was considered to the more complicated case of multiplicative disturbances. The obtained results make it possible to evaluate disturbances caused by uncertain or perturbed parameters of dynamical systems (e.g., stiffness, damping factors, feedback coefficients, electrical parameters, etc.).

Keywords: Dynamical System, Reachable Set, Uncertain Parameters, State Estimation.

Petrosjan. L. A., 1997, "The Time-consistency Problem in Nonlinear Dynamics" , RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 291-303.

The nonlinear dynamic optimization problem (NDOP) $\Gamma(x0,[t0,T])$ on the time interval $[t_0,T]$ from the initial state x0 is considered. If the NDOP is multicriterial or game-theoretic, there exists a rich variety of optimality principles each one consisting from a set of optimal decisions. Denote one such optimality principle in $\Gamma(x_0,[t_0,T])$ by $M(x_0,[t_0,T])$. For a given optimal decision $\overline{m}\,(x_0,[t_0,T]) \in M\,(x_0,[t_0,T])$ let $\bar{x}\,(t)$ be the corresponding optimal trajectory. Consider NDOP subproblems along $\bar{x}\,(t)\Gamma(\bar{x}\,(t),[t,\ T]),\ t\in,[t_0,T]$. Let $M(\bar{x}\,(t),\ [t,\ T])$ be the corresponding optimality principle. Denote by $\overline{m}\,(\bar{x}\,(t),[t,T])$ the trace of optimal decision $\overline{m}\,(x_0,[t_0,T])$ in subproblem $\Gamma(\bar{x}\,(t),[t,\ T])$. The optimality principle is called time-consistent if for every $m\,(x_0,[t_0,T]) \in M\,(x_0,[t_0,T])$ the trace $\overline{m}\ (\bar{x}\,(t),\ [t,\ T]) \in M\,(\bar{x}\,(t),[t,T])$ Most of the classical optimality principles in NDOP $\Gamma(x0,[t0,T])$ (as Shown in (Petrosjan, 1993; Petrosjan and Zenkevich, 1996)) are time inconsistent. The "agreeable solution" (Kaitala and Pohjola, 1988) as an optimality principle in NDOP is considered its time-consistency investigated and the regularization procedure which guarantees the time-consistency of the regularized optimality principle proposed.

Keywords: Agreeable Solution, Nash Equilibria, Pareto Optimality, Time-Consistency.

Ivanovych, I. V., and Volodymyrovych, Z. I., 1997, "About Some Properties of the Statistical Regularities", RBCM - J. of the Braz. Soc. Mechanical Sciences, Vol. 19, No. 2, pp. 304-312.

In this article the notion of G_f-independence of two sequences has been formulated in the terms of their joint statistical regularity. The example which illustrates the construction of the statistical regularity of the sequence, taking values from the finite set, has been given.

Keywords: Statistical Regularity, G_f-Independence of Two Sequences, Finitely Additive Measure.

Note for Contributors: Articles on Disk

- Authors are strongly encouraged to submit the final accepted manuscripts on disk, using text editors for Windows or Word for Windows.
- The disk must be marked with the paper identification number and software used. Two copies of the printout should be included.

Sixth Pan American Congress of Applied Mechanics
PACAM VI

to be held jointly with

8th International Conference on Dynamic Problems in Mechanics
DINAME 99
in
Rio de Janeiro
Brazil
4 - 8 January, 1999

PACAM I was held in 1989 in Rio de Janeiro. After ten years of successful meetings (Rio de Janeiro - BRAZIL, 1989; Valparaiso - CHILE, 1991; São Paulo - BRAZIL, 1993; Buenos Aires - ARGENTINA, 1995; San Juan - PUERTO RICO, 1997), we are pleased to announce that PACAM will return to Rio de Janeiro.

PACAM aims to bring together researchers, practicing engineers and students from South, Central, and North America. However, participants from all other continents have taken part in previous meetings and are welcome to PACAM VI. An unusual opportunity is provided for personal interaction between workers from different geographical areas and from different branches of mechanics.

Papers on all the usual, as well as unusual, topics of mechanics are welcome. We wish to highlight the social development: "Women in Mechanics" and the technical development: "Biomechanics".

Sponsors:

The American Academy of Mechanics
The Brazilian Academy of Engineering
The Brazilian Society of Mechanical Sciences

Technical Sessions:

Pontificia Universidade Católica
Rua Marques de São Vicente 225 Gávea
Rio de Janeiro, RJ Brasil

Accommodations:

Copacabana Beach or Ipanema Beach

Deadlines:

April 1, 1998 - Submission of 4-page abstract for the Congress.
April 1, 1998 - Application for travel grant for participants from the U.S. (Funding pending)
August 1, 1998 - Notification of acceptance of paper.

More information: pacam99@civ.puc.rio.br

RBCM - J. of the Braz. Soc. Mechanical Sciences
Vol. XIX - No.2 - 1997 - pp. 318-321

ISSN 0100-7386
Printed in Brazil

FORMULÁRIO DE AFILIAÇÃO

ASSOCIAÇÃO BRASILEIRA DE CIÊNCIAS MECÂNICAS
Av. Rio Branco, 124 - 18º andar - 20040-001 Rio de Janeiro - RJ
Tel.: (021) 221-0438 - Fax.: (021) 222-7128
e-mail: abcmalfs@ omega.lncc.br
CGC 83.431.593/0001-78

INDIVIDUAL

Por favor, preencha os dois lados do formulário.

Nome _____

Endereço Residencial _____

CEP _____ Cidade _____ Estado _____

País _____ Tel.: ( ) _____ Fax: ( ) _____

E-mail _____

Empresa _____

Dept./Divisão _____ Posição _____

Endereço Comercial _____

CEP _____ Cidade _____ Estado _____

País _____ Tel.: ( ) _____ Fax: ( ) _____

E-mail _____

Candidato-me a: ☐ Admissão ☐ Mudança de Categoria

Na categoria de: ☐ Sócio efetivo ☐ Sócio Aspirante

Solicito enviar correspondência para o seguinte endereço:

☐ Comercial ☐ Residencial

Data _____ Assinatura _____

Para uso da ABCM

Aprovado _____ Data _____ Sócio nº _____

FORMAÇÃO E EXPERIÊNCIA PROFISSIONAL

Por favor, liste em ordem cronológica os dados completos de sua formação e experiência profissional. A falta desses dados impedirá o processo de admissão. Obrigado.

FORMAÇÃO ACADÊMICA

Graduação - Área _____ Anos _____ a _____
Instituição _____ País _____

Mestrado - Área _____ Anos _____ a _____
Instituição _____ País _____

Doutorado - Área _____ Anos _____ a _____
Instituição _____ País _____

Outro - Área _____ Anos _____ a _____
Instituição _____ País _____

EXPERIÊNCIA PROFISSIONAL

Empresa _____ Anos _____ a _____
Natureza da atividade _____ Posição _____

Empresa _____ Anos _____ a _____
Natureza da atividade _____ Posição _____

Empresa _____ Anos _____ a _____
Natureza da atividade _____ Posição _____

Indique até um máximo de 8 áreas de acordo com os códigos numéricos do Anexo.

Áreas de Especialização _____

Aplicação _____

Comentários _____

Área de Especialização

Especifique no Formulário de Afiliação os códigos numéricos das suas Áreas de Especialização e de Aplicação (verso).

1000 Fundamentos e Métodos Básicos em Mecânica Teórica e Aplicada
1010 Mecânica do Contínuo
1110 Método dos Elementos Finitos
1120 Método dos Elementos de Contorno
1130 Métodos Assintóticos
1140 Método das Diferenças Finitas
1150 Outro Métodos em Mec. Computacional
1210 Métodos Estocásticos e Estatísticos
1310 Modelagem
1410 Fundamentos de Análise Experimental
1510 Metrologia
1610 Gerência de Projetos
2000 Dinâmica e Vibrações
2110 Cinemática e Dinâmica
2210 Vibrações de Sólidos - Fundamentos
2310 Vibrações - Elementos de Estruturas
2320 Vibrações - Estruturas
2330 Propagação de Ondas em Sólidos
2340 Impacto em Sólidos
2350 Identificação de Parâmetros
2420 Propagação de Ondas em Fluídos
2510 Interação Fluido-Estrutura
2610 Astronáutica - Mec. Celeste e Orbital
2710 Explosão e Balística
2810 Acústica
3000 Controle e Otimização
3110 Projeto e Teoria de Sistemas Mecânicos
3210 Sistemas de Controle Ótimo
3220 Sistemas de Controle Adaptativo
3230 Aplicações em Sistemas e Controle
3310 Robótica
3410 Otimização de Sistemas e Processos
4000 Materiais
4110 Biomateriais
4120 Materiais Metálicos
4130 Materiais Cerâmicos
4140 Materiais Poliméricos
4150 Materiais Conjugados
4210 Conformação Mecânica
4300 Caracterização e Controle Microestrutural
4410 Comp. Mecânico dos Materiais
4420 Comp. Mec. Mat. - Baixas Temperaturas
4430 Comp. Mec. Mat. - Altas Temperaturas
4440 Comp. Mec. Mat. - Carregamto. Variável
4450 Comp. Mec. Mat. - Carregamto. Dinâm.
4500 Mecanismos de Fratura
4600 Mecânica da Fratura
4710 Ensaios Destrutivos
4720 Ensaios Não-Destrutivos
4800 Corrosão
5000 Mecânica dos Sólidos
5010 Elasticidade Linear
5020 Elasticidade Não-Linear
5030 Viscoelasticidade
5040 Plasticidade
5050 Visco-Plasticidade
5060 Mecânica dos Materiais Conjugados
5070 Mecânica dos Meios Porosos
5110 Reologia
5210 Cabos, Hastes e Vigas
5220 Membranas, Placas e Cascas
5230 Estruturas - Geral
5240 Estruturas - Contato com o Solo
5250 Estruturas - Submersas/Semi-submersas
5260 Estruturas - Móveis
5270 Estruturas - Vasos e Contenção
5310 Mecânica dos Solos - Básico
5320 Mecânica dos Solos - Aplicações
5330 Mecânica das Rochas
5410 Efeitos Eletro-Magnéticos em M. Sólidos
5420 Efeitos Térmicos em M. dos Sólidos
5510 Estabilidade de Estruturas
5520 Comportamento após a Flambagem
5530 Estados Limites e Cargas de Colapso
5540 Acomodação e Acúmulo de Dano
5610 Mecânica de Fratura
5650 Tribologia
5655 Atrito e Desgaste
5710 Componentes de Máquinas
5720 Acoplamentos e Juntas Não-Soldadas
5800 Análise Experimental de Tensões
6000 Mecânica dos Fluidos
6010 Reologia
6110 Hidráulica
6210 Escoamento Incompressível
6220 Escoamento Compressível
6230 Escoamento Rarefeito
6240 Escoamento em Meios Porosos
6250 Magneto-Hidrodinâmica e Plasma
6270 Escoamento Multifásico
6310 Camada Limite - Contorno Sólido
6320 Camada Limite - Contorno Livre
6410 Escoamento Interno - Dutos, Canais, etc.
6430 Escoamento com Superfície Livre
6510 Estabilidade de Escoamento
6520 Turbulência
6530 Hidrodinâmica - Veículo de Estrut. Naval
6540 Aerodinâmica
6610 Mec. Fluídos - Aplicações em Máquinas
6650 Lubrificação
6710 Transientes em Fluidos
6810 Téc. Experim. e Visualização Escoamento
7000 Termociências
7010 Termodinâmica
7110 Transp. de Calor - Convec. Monofásica
7120 Transp. de Calor - Convec. Bifásica
7130 Transp. de Calor - Condução
7140 Transp. de Calor - Radiação/Mod. Comb
7150 Transp. de Calor - Dispositivos/Sistemas
7210 Termodinâmica dos Sólidos
7310 Transporte de Massa
7410 Combustão
7420 Combustão em Leito Fluidizado
7510 Acionadores e Dispositivos de Propulsão
8000 Geociências
8010 Micromeritics
8110 Meios Porosos
8210 Geomecânica
8310 Mecânica dos Abalos Sísmicos
8410 Hidrologia, Oceanografia, Meteorologia
9000 Energia e Meio Ambiente
9110 Combustíveis Fósseis
9120 Sistemas Nucleares - Fissão
9125 Sistemas Nucleares - Fusão
9130 Sistemas Geotérmicos
9140 Sistemas Solares
9150 Sistemas Eólicos
9160 Sistemas de Energia Oceânica
9210 Armazenamento de Energia
9220 Distribuição de Energia
9310 Mecânica dos Fluidos Ambiental
9410 Mecânica de Dispositivos de Armazenamento de Resíduo
10000 Biociências
10110 Biomecânica
10210 Ergonomia
10310 Reabilitação
10410 Mecânica nos Esportes

Áreas de Aplicação

Exemplo: um especialista em Mecânica dos Fluídos (família 6000) atuando na área de Turbulência (6520), deverá escolher a Área de Aplicação 350, se estiver trabalhando em Propulsão.

010 Acústica e Controle de Ruído
020 Aplicações em Biociências
030 CAD
040 CAM
050 Componentes de Máquinas
060 Controle Ambiental
070 Controle de Qualidade
080 Criogenia
090 Engenharia e Física de Reatores

100 Engenharia de Petróleo
110 Engenharia Oceanográfica
120 Equipamentos de Processos
130 Equipamentos Industriais
140 Fontes Alternativas de Energia
150 Forjamento
160 Fundição
170 Garantia de Qualidade
180 Indústria Têxtil e Tecnologia Correlata
190 Inspeção e Certificação

200 Instalações Industriais
210 Instrumentação
220 Lubrificação Industrial
230 Mancais e Rolamentos
240 Máquinas Ferramentas
250 Máquinas de Fluxo
260 Máquinas Motrizes
270 Mecânica Fina
280 Metalurgia Geral e Beneficiamento de Minério
290 Metrologia

300 Mineração e Metalurgia Extrativa
310 Óptica
320 Pontes e Barragens
330 Processos de Fabricação
340 Projeto de Estruturas
350 Propulsão
360 Prospecção e Propulsão
370 Servo Mecanismos e Controle
380 Siderurgia
390 Sistemas Hidráulicos

400 Sistemas Pneumáticos
410 Soldagem
420 Solicitações Acidentais - Efeitos de Vento, Sismo, Explosão, Fogo e Inundação
430 Tecnologia de Alimentos
440 Tecnologia Mineral
450 Transporte (excluído veículos)
460 Transmissão de Energia
470 Tratamento Térmico e Termoquímico
480 Tubulações Industriais e Nucleares
490 Usinas Hidrelétricas

500 Usinas Termoelétricas
510 Vácuo
520 Vasos de Pressão, Trocadores de Calor e Equipamentos Pesados
530 Veículos - Terrestres, Espaciais e Marítimos

Journal of the Brazilian Society of Mechanical Sciences

Information for Authors

SCOPE AND POLICY

- The purpose of the Journal of the Brazilian Society of Mechanical Sciences is to publish papers of permanent interest dealing with research, development and design related to science and technology in Mechanical Engineering, encompassing interfaces with Civil, Electrical, Chemical, Naval, Nuclear, Agricultural, Materials, Petroleum, Aerospace, Food, System Engineering, etc., as well as with Physics and Applied Mathematics.
- The Journal publishes Full-Length Papers, Review Papers and Letters to the Editor. Authors must agree not publish elsewhere a paper submitted to and accepted by the Journal. Exception can be made for papers previously published in proceedings of conferences. In this case it should be cited as a footnote on the title page. Copies of the conference referees reviews should be also included. Review articles should constitute a critical appraisal of published information.
- The decision of acceptance for publication lies with the Editors and is based on the recommendations of at least two ad hoc reviewers, and of the Editorial Board, if necessary.

SUBMISSION

- Manuscripts and all the correspondence should be sent to the Editor or, alternatively, to the appropriate Associate Editor.
- Four (4) copies of the manuscript are required. The author should submit the original figures, which will be returned if the paper is not accepted after the review process.
- Manuscripts should be submitted in English or Portuguese. Spanish will also be considered.
- A manuscript submitted for publication should be accompanied by a cover letter containing the full name(s) of author(s), mailing adresses, the author for contact, including phone and fax number, and, if the authors so wish, the names of up to five persons who could act as referees.

FORMAT

- Manuscripts should begin with the title, including the english title, the abstract and up to five key words. If the paper's language is not English, an extended summary of about 500 words should be included. The manuscript should not contain the author(s) name(s).
- In research papers, sufficient information should be provided in the text or by refering to papers in generally available Journals to permit the work to be repeated.
- Manuscripts should be typed double-spaced, on one side of the page, using A-4 sized paper, with 2 cm margins. The pages should be numbered and not to excess 24 pages, including tables and figures. The lead author of a RBCM paper which exceeds the standard length of pages will be assessed a excess page charge.
- All symbols should be defined in the text. A separate nomenclature section should list, in alphabetical order, the symbols used in the text and their definitions. The greek symbols follow the English symbols, and are followed by the subscripts and superscripts. Each dimensional symbol must have SI (Metric) units mentioned at the end. In addition, English units may be included parenthetically. Dimensionless groups and coefficients must be so indicated as dimensionless after their definition.
- Uncertainties should be specified for experimental and numerical results.

ILLUSTRATIONS AND TABLES

- Figures and Tables should be referred in consecutive arabic numerals. They should have a caption and be placed as close as possible to the text first reference.
- Line drawings should be prepared on tracing paper or vellum, using India ink; line work must be even and black. Laser print output is acceptable. The drawings with technical data/results should have a boundary on all four sides with scale indicators (tick marks) on all four sides. The legend for the data symbols should be put in the figure as well as labels for each curve wherever possible.
- Illustrations should not be larger than 12 x 17 cm. Lettering should be large enough to be clearly legible (1.5-2.0 mm).
- Photographs must be glossy prints.

REFERENCES

- References should be cited in the text by giving the last name of the author(s) and the year of publication of the reference: either "Recent work (Smith and Jones, 1985) ... or "Recently Smith and Jones (1985) With four or more names, use the form "Smith et al.(1985)" in the text. When two or more references would have the same text identification, distinguish them by appending "a", "b", etc., to the year of publication.
- Acceptable references include: journal articles,dissertations, published conference proceedings, numbered paper preprints from conferences, books, submitted articles if the journal is identified, and private communications.
- References should be listed in alphabetical order, according to the last name of the first author, at the end of paper. Some sample references follow:

Bordalo, S.N., Ferziger, J.H. and Kline, S.J., 1989, "The Development of Zonal Models for Turbulence", Proceedings, 10th ABCM - Mechanical Engineering Conference, Vol. I, Rio de Janeiro, Brazil, pp. 41-44.

Clark, J.A., 1986, Private Communication, University of Michigan, Ann Arbor, MI.

Coimbra, A.L., 1978, "Lessons of Continuum Mechanics", Editora Edgard Blucher Ltda, São Paulo, Brazil.

Kandlikar. S.G. and Shah, R.K., 1989, "Asymptotic Effectiveness - NTU Formulas for Multiphase Plate Heat Exchangers", ASME Journal of Heat Transfer, Vol. 111, pp. 314-321.

McCormak, R.W., 1988, "On the Development of Efficient Algorithms for Three Dimensional Fluid Flow", Journal of The Brazilian Society of Mechanical Sciences, Vol. 10, pp. 323-346.

Silva, L.H.M., 1988. "New Integral Formulation for Problems in Mechanics", (in portuguese), Ph.D. Thesis, Federal University of Santa Catarina, Florianópolis, SC, Brazil.

Sparrow, E.M., 1980a, "Forced-Convection Heat Transfer in a Duct Having Spanwise-Periodic Rectangular Protuberances", Numerical Heat Transfer, Vol. 3, pp. 149-167.

Sparrow, E.M., 1980b, "Fluid-to-Fluid Conjugate Heat Transfer for a Vertical Pipe-Internal Forced Convection and External Natural Convection", ASME Journal of Heat Transfer, Vol. 102, pp. 402-407.

ISSN 0100-7386

JOURNAL OF THE BRAZILIAN SOCIETY OF MECHANICAL SCIENCES
REVISTA BRASILEIRA DE CIÊNCIAS MECÂNICAS

VOL. XIX - No. 2 - JUNE 1997

DYNCO - NINTH WORKSHOP ON DYNAMICS AND CONTROL

Rio de Janeiro, RJ-August 11-14, 1996

Contents continued on page 312