Estimation of non-linear dynamical systems with application to landing gear

WFW 91.016

Korte Onderzoekersopleiding Fundamentele Werktuigkunde

ESTIMATION OF NON–LINEAR DYNAMICAL SYSTEMS WITH APPLICATION TO LANDING GEAR

G. Verbeek

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Verbeek, G.

Estimation of non-linear dynamical systems with application to landing gear / G. Verbeek. - Eindhoven : Instituut Vervolgopleidingen, Technische Universiteit Eindhoven. - Ill.
Korte Onderzoeksopleiding Fundamentele Werktuigkunde. - Met lit. opg.
ISBN 90-5282-095-3 geb.
Trefw.: systeem identificatie bij landingsgestellen / periodieke oplossingen bij landingsgestellen.

TABLE OF CONTENTS

Summary v

Notation vii

List of symbols viii

0. General introduction 1

1. Introduction 3

2. Problem definition 5

3. Model estimation 11

4. Optimization 21

5. Dynamical systems 27

5.1 Time integration method 27
5.2 Periodic solution method 31

6. Parameter estimation of an oleo 35

6.1 Time integration method 41
6.2 Periodic solution method 44

7. Conclusions 51

Literature 53

Appendix 55

SUMMARY

In this report the parameter estimation of non – linear dynamical systems and application to aircraft landing gear dampers, often called oleo's, is discussed. It appears that this estimation problem can be solved by application of two different solution techniques to obtain the dynamical responses, namely by time integration of the state equations or by calculating a periodic solution due to a periodic excitation signal. These solution techniques are discussed in sections 5.1 and 5.2 respectively. The application of both methods to an oleo is discussed in chapter 6. The periodic estimation method gives the best results. This method will be employed in a new test method for aircraft landing gear.

The preceding chapters contain the necessary theoretical basis of the estimation process. In chapter 2 the differences between deterministic and probabilistic models are discussed. It is shown that the optimal value of a probabilistic model can be defined in different ways. In the strategy selected, the optimal value generally is the solution of an optimization problem. Chapter 3 displays the differences between the objective functions of four frequently used probabilistic models: Bayesian estimation, maximum likelihood estimation, weighted least squares or Markov estimation and unweighted least squares estimation. The fourth chapter deals with non – linear optimization methods needed to optimize the derived objective functions. The report is completed with conclusions in chapter 7. It appears that the estimation problem can be solved most conveniently by maximum likelihood estimation with normal distributed errors and prior information concerning the parameters. The resulting numerical problem is solved most robustly by application of a modified Gauss – Newton technique.

NOTATION

a, A	scalars, or scalar functions
$\mathbf{a}$	column with scalars or functions as components
$\mathbf{A}$	matrix with scalars or functions as components
a_b	subscripts
$a_{,b}$	partial derivative of a(b,c) with respect to b . The variables can either be scalars, vectors or matrices as long as the result is an array of up to two dimensions, otherwise summation notation will be used
$\mathbf{a}^T, \mathbf{A}^T$	transpose of a column and matrix respectively
$\mathbf{A}^{-1}$	inverse of a matrix
$\mathbf{A}^+$	pseudo – inverse of a matrix
$\dot{\mathbf{a}}$	total derivative of $\mathbf{a}$ with respect to time t
$\mathbf{a}'$	total derivative of $\mathbf{a}$ with respect to cyclic time τ
$\hat{\mathbf{a}}$	exact or true value of $\mathbf{a}$
$\bar{\mathbf{a}}$	mean value of $\mathbf{a}$
$\mathbf{a}^*$	posterior or estimated value of $\mathbf{a}$
$\mathbf{a}_0$	prior or initial value of $\mathbf{a}$

LIST OF SYMBOLS

a	estimator function
b	bias
d	step size in discretization process
e	residual vector
f	frequency
f	output function vector in dynamical models
g	static and dynamical deterministic model function vector
h	state equation vector
i	index for the number of replications of an experiment
j	current time index
k	size of the total variable vector **w**
ℓ	size of the set of parameters $\boldsymbol{\theta}$
m	size of the set of model functions **g**
n	size of the set of experiments
p	probability density function
p1	first partition of $\ell-$ sized variables
p2	second partition of $\ell-$ sized variables
q	gradient of an objective function ϕ
r	size of the set of independent variables **x**
s	size of the set of dependent variables or displacements
s	state variable vector
t	time
u	displacement vector
w	total variable vector in static models
x	independent variable vector
y	dependent (output, observed) variable vector
z	periodic solution vector of a dynamical system

C	constants
$\mathbf{E}$	matrix form of the error vector $\mathbf{e}$
E	expected value operator
$\mathbf{H}$	hessian matrix
$\mathbf{J}$	jacobian matrix
L	likelihood function
$\mathbf{M}$	moment matrix of the residuals
N	normal distribution, number of replications of the experiments
$\mathbf{Q}$	second order information in the hessian matrix
$\mathbf{S}$	matrix of the singular value decomposition
$\mathbf{U}$	matrix of the singular value decomposition
$\mathbf{V}$	covariance matrix, matrix of the singular value decomposition
$\mathbf{W}$	weight factor matrix
$\mathbf{Y}$	matrix form of the vector $\mathbf{y}$
α	index for the parameter vector $\boldsymbol{\theta}$
β	index for the observed variable vector $\mathbf{y}$
δ	variation operator
ϵ	error vector
$\boldsymbol{\theta}$	parameter vector
λ	scalar used in the Levenberg – Marquardt minimization
μ	index used to label the experiments
σ	variance
τ	cyclic time
$\boldsymbol{\phi}$	augmented parameter vector
$\boldsymbol{\psi}$	vector form of the probability distribution parameter matrix
Δ	scalar used in the Levenberg – Marquardt minimization
ϕ	objective function
Ψ	probability distribution parameter matrix

0. GENERAL INTRODUCTION

Nowadays when we speak of an aircraft landing gear we have in view a mechanical system consisting of several subsystems as there are: single or multiple tyres, brakes, structural parts, hydraulic retraction and locking devices, the dissipative oleo/pneumatic element and the necessary control systems. These landing gears have been developed from very simple rigid gears. In the early days of aviation, for example, the soft tyres themselves were the main dissipative parts of the landing gear. A review of historic and present landing gear designs can be found in an article by Young (1986).

From the beginning of this century to World War II there has been some progress in landing gear design but hardly any in landing gear dynamics. This was mainly due to the capability of the existing landing gears to handle the landing loads of the aircrafts of that era. As World War II strongly influenced several technical evolutions, also the rate of progress in landing gear design and dynamics rapidly increased during and after the war. The demand for faster and heavier military aircraft made it necessary to do research on landing gear dynamics. Of primary concern was the dissipation of the horizontal and vertical kinetic energy during the impact and the roll–out phase of the landing. From this period Milwitzky and Cook's (1953) report is still a very often referenced work and describes the basic nonlinear analytical model used in many modern landing simulation codes. Socalled 'Drop Tests' and 'Landing Tests' were, and still are, used for the validation of these nonlinear dynamical models.

In the period following Milwitzky and Cook's report hardly any numerical or experimental studies were reported in which the objective was to improve the nonlinear analytical landing gear model. Instead of developing more accurate gear models, emphasis was placed on the numerical simulation of ground operations as taxiing, cornering and towing. Ground operation loads could be of greater importance than the impact landing loads, for example fatigue loads due to the extended operating life for civil aircraft (taxi distance $\approx$ 500.000 km) or roll–out and lift–off loads for military aircraft on bomb damage repaired or emergency runways. For validation of these numerical simulations of ground operations other test procedures are being used, such as 'Linear Shaker Tests' and 'Taxi Tests'.

Due to the high investments for a drop test facility and expected low applicability for the design process or optimization, DAF Special Products initialised research for an alternative 'Shaker Test' procedure for aircraft landing gear. A literature study (Verbeek 1989) revealed the necessity for improved nonlinear dynamical models of landing gear that should be validated by suitable identification methods. The object of research has been constrained to modelling and identification of aircraft landing gear dampers, often called oleo's, because of the uncertainties in the physical nonlinearities of the oleo. In contrast, the geometrical nonlinearities of the complete gear can readily be modelled and simulated by multi – body dynamics codes, so this need not be a topic of further research. The final aim of the research is the assessment of the new experimental shaker test method, that is illustrated in figure 1.

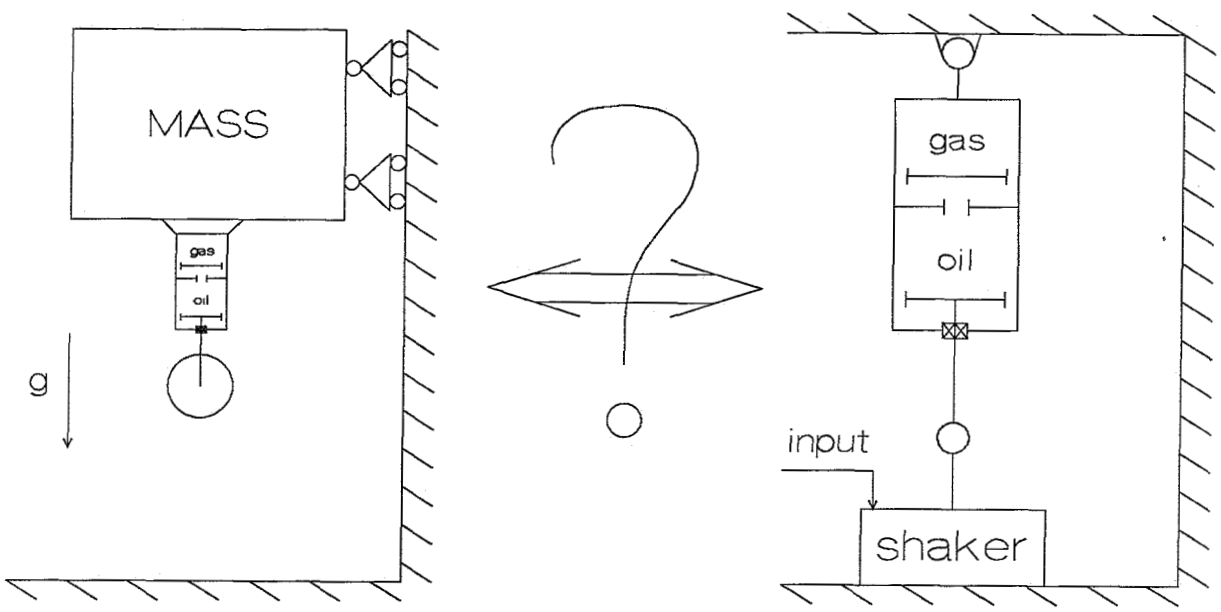

Figure 1. Objective of research, drop test versus alternative shaker test

The research is performed in cooperation with the section of Engineering Fundamentals of the department of Mechanical Engineering, Eindhoven University of Technology.

1. INTRODUCTION

During the research on modelling techniques for aircraft landing gear and especially landing gear dampers, often called oleo's, it appeared necessary to perform a study on parameter estimation theory. In literature only one monograph by Batill & Bacarro (1988) was found that stressed the need of accurate parameter estimation procedures for the experimental verification of assumed oleo models. The estimator used in this study was the well known nonlinear unweighted least squares estimator, which makes use of Newton–Gauss iteration. Many problems were encountered such as: divergence by ill–conditioning, the sensibility to scaling of the variables, numerical overflow by inadmissable parameter combinations or discontinuities in friction functions. In figure 2 it can be seen how parameter estimation techniques can be helpful to improve the parameters in an assumed model for an oleo, landing gear or in principal any other dynamical system.

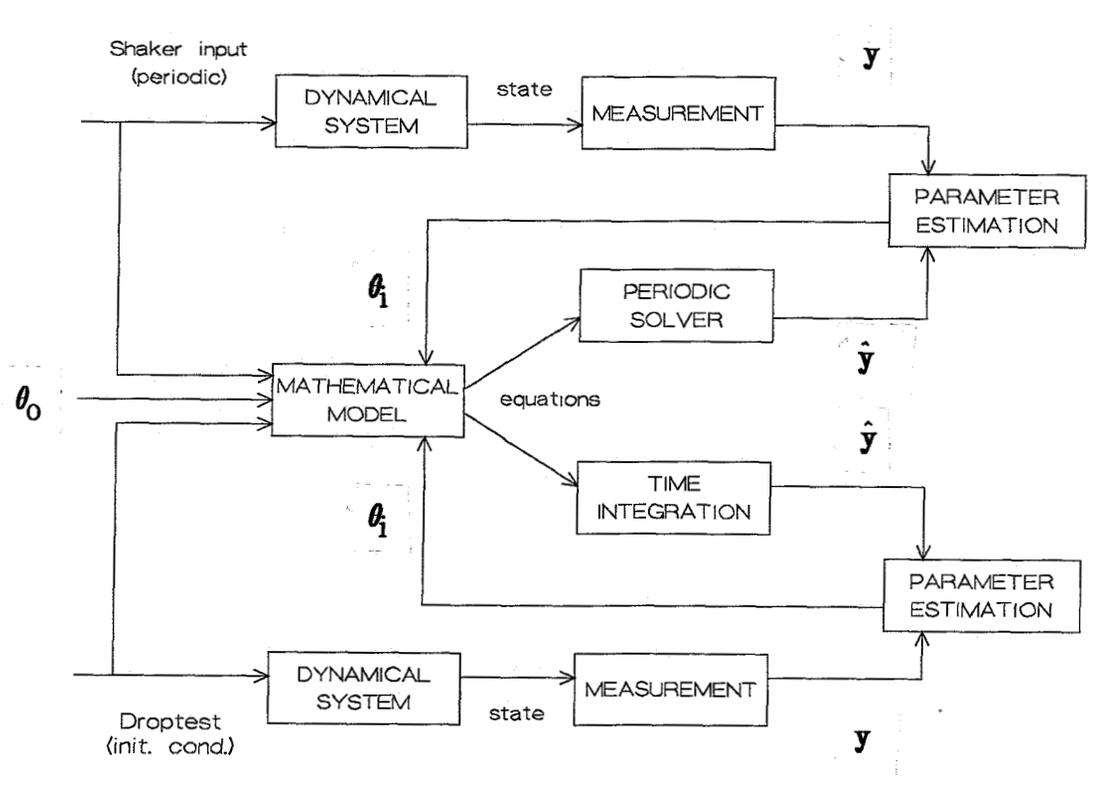

figure 2. Drop test and alternative shaker test experiments and computations.

In the lower branch the conventional drop test procedure is illustrated. Starting with the drop test initial conditions, the drop test can be performed and the state vector or some of it's components or output signals can be measured. The initial conditions together with approximate parameters can be used in an assumed mathematical model and by time integration procedures the measured respons can be predicted. Based on the discrepancy between the measured and predicted output, parameter estimation procedures can be used to improve the approximate parameter vector.

The upper branch is an illustration of the proposed alternative shaker test procedure. Input to the system and the assumed mathematical model is now a periodic signal. The predicted output components, in this case, can be calculated by a periodic solver which will drop the need to estimate the initial conditions. Analogously to the procedure mentioned above the improved parameter vector can be computed by an appropriate parameter estimation technique.

In the article by Batill a *model fitting* procedure was used. The output of this procedure is a so called point estimate (e.g. $\theta = 4$). Only the parameter values are computed and nothing is said about the confidence in the computed values. This is merely a form of data reduction with which the measured response can be recomputed or hopefully predicted. Nothing can be said about the validity of the mathematical model. Application of estimation theory will lead to *model estimation* by taking into account the stochastic properties of the measurements. This will lead to interval estimates (e.g. $\theta = 4 \pm 0.2$).

Another main issue that will be dealt with in the sequel is the difference in estimating static and dynamical models. The most often used estimation theory deals with the estimation of parameters in (nonlinear) algebraic equations. It can be found in most good textbooks on estimation theory. The estimation of dynamical systems is more complicated because the parameters can occur not only in the differential equations but also in the initial conditions and the output equations too.

2. PROBLEM DEFINITION

Deterministic models

In parameter estimation problems the deterministic mathematical model usually consists of nonlinear algebraic equations of the following general functional form

$$\mathbf{g}(\mathbf{w},\boldsymbol{\theta}) = \mathbf{0} \quad , \tag{2.1}$$

in which $\mathbf{g}$ is a m – dimensional vector function, $\mathbf{w}$ is a k – dimensional column of variables and $\boldsymbol{\theta}$ is the unknown ℓ – dimensional parameter vector. With this deterministic model it is only possible to compute the unknown parameters from measured variables $\mathbf{w}$ that are assumed to be exact. Often these models appear in reduced form for prediction purposes

$$\mathbf{y} = \mathbf{f}(\mathbf{x},\boldsymbol{\theta}) \quad , \tag{2.2}$$

in which $\mathbf{w}$ is divided in the independent r – dimensional variable vector $\mathbf{x}$ and the s – dimensional dependent or observed variable vector $\mathbf{y}$.

Experimental data

Before we are able to calculate or estimate any parameters of a model, we will have to define and arrange a set of experiments. By performing the experiments one stores data of the observed variables under certain experimental conditions. This will lead to an immens amount of data that has to be stored for off – line calculations. The experiments can be labeled by application of an index μ ($\mu = 1,2, \ldots ,n$). The observed values, and lateron all analogously derived variables, can be stored in vectors $\mathbf{y}_\mu$, in a matrix

$$\mathbf{Y} = [\mathbf{y}_1;\mathbf{y}_2; \ldots ;\mathbf{y}_n]^T \quad , \tag{2.3}$$

or in a vector

$$\mathbf{y} = [\mathbf{y}_1{}^T, \mathbf{y}_2{}^T, \ldots, \mathbf{y}_3{}^T]^T \quad . \tag{2.4}$$

Probabilistic models

As pointed out above the deterministic models will not be applicable when, for instance, the measurements are contaminated with random noise or when the mathematical model not exactly represents reality. These imperfections can be imbedded in the model by probability statements on the errors.

The transition from deterministic models to probabilistic models is discussed in Bard (1974). Here only the probabilistic form of the reduced model (2.2) with exact measured independent variables $\mathbf{x}$ will be discussed, because of all models the estimation of parameters for this model will be easiest. On the other hand if the errors in $\mathbf{x}$ were known, it would be very difficult to predict the effect in the resulting errors.

So the lack of fit of the data to the deterministic model can only be of two different origins: errors due to the observed variables and errors due to the model equations. We will not distinguish the two errors and so the sum can be written as one total error ϵ_μ . Addition to the deterministic model (2.2) gives

$$\mathbf{y}_\mu = \mathbf{f}(\mathbf{x}_\mu, \boldsymbol{\theta}) + \epsilon_\mu \quad , \tag{2.5}$$

in which $\boldsymbol{\theta}$ denotes the true parameter values. We also can define the exact or underlying measurements as $\hat{\mathbf{y}}_\mu \equiv \mathbf{y}_\mu - \epsilon_\mu$ and write the model as

$$\hat{\mathbf{y}}_\mu = \mathbf{f}(\mathbf{x}_\mu, \boldsymbol{\theta}) \quad . \tag{2.6}$$

The error matrix $\mathbf{E}$ (definition analogue to the experimental data matrix) can be thought of as a realization of a stochastic matrix variable. The so – called joint probability density function that belongs to the errors can be written as

$$p(\mathbf{E}|\Psi) = \prod_{\mu=1}^{n} p(\epsilon_\mu|\Psi_\mu) \quad , \tag{2.7}$$

in which Ψ represents the given parameters of the distribution functions. Such a joint

probability density function can be illustrated by an example. Let the errors of the experiments be normally distributed with mean zero and covariance matrix $\mathbf{V}_\mu$ for all s observed variables $\mathbf{y}_\mu$, thus $N_s(0,\mathbf{V}_\mu)$ and independent in different experiments, then the following joint probability density function can be derived

$$p(\mathbf{E}|\mathbf{V}_1,\mathbf{V}_2,\ldots,\mathbf{V}_n) =$$

$$(2\pi)^{-ns/2} \prod_{\mu=1}^{n} (\det^{-1/2} \mathbf{V}_\mu) \exp\left[-\frac{1}{2}\sum_{\mu=1}^{n} \epsilon_\mu{}^T \mathbf{V}_\mu^{-1} \epsilon_\mu\right] \quad . \tag{2.8}$$

This function gives us all the stochastic information on the newly added error ϵ_μ to complete the probabilistic model. For estimation purposes from this function we can derive the likelihood function. We simply have to substitute the expressions of the residuals for the errors in the last formula. The residuals are defined by the same formula as the errors but result from a calculation with an inexact or estimated parameter vector $\boldsymbol{\theta}$

$$\mathbf{e}_\mu = \mathbf{y}_\mu - \mathbf{f}(\mathbf{x}_\mu,\boldsymbol{\theta}) \quad . \tag{2.9}$$

The resulting likelihood function reads

$$L(\boldsymbol{\theta}|\mathbf{V}_1,\mathbf{V}_2,\ldots,\mathbf{V}_n) = (2\pi)^{-ns/2} \prod_{\mu=1}^{n} (\det^{-1/2} \mathbf{V}_\mu) \times$$

$$\times \exp\left[-\frac{1}{2}\sum_{\mu=1}^{n} [\mathbf{y}_\mu - \mathbf{f}(\mathbf{x}_\mu,\boldsymbol{\theta})]^T \mathbf{V}_\mu^{-1} [\mathbf{y}_\mu - \mathbf{f}(\mathbf{x}_\mu,\boldsymbol{\theta})]\right] \quad . \tag{2.10}$$

The difference in the model fitting and the model estimation problem should now be clear. We not only have to estimate the parameters $\boldsymbol{\theta}$ but also the distribution parameters Ψ , or in the normal distributed case the covariance matrices $\mathbf{V}_\mu$. The augmented parameter vector will be denoted $\boldsymbol{\phi} = [\boldsymbol{\theta}^T;\boldsymbol{\psi}^T]^T$ (with matrix Ψ transformed to a vector form $\boldsymbol{\psi}$ according to equation (2.3)). For other models the vector can even be extended with the exact values of the measurements $\hat{\mathbf{y}}$. We will avoid this for mathematical simplicity and practical considerations. Adding $\hat{\mathbf{y}}$ would introduce an additional large number of variables. An application of such a technique can be found in Molengraft (1990). Essentially it is a least squares technique for estimation of the unknown parameters $\boldsymbol{\theta}$ extended with the estimation of the exact measurements. The distribution parameters are

predefined by weight matrices. Usually assumptions on the probability distributions are made by the users of estimation procedures implicitly. The weights can be estimated as unknown parameters of the probability distributions instead of assigning them.

Prior information

We still can add more information to the estimation problem, which can be helpful in the solution procedure. One usually has some knowledge on the range or bounds of the admissable parameters. This is called prior information to the estimation process. The prior distribution of the parameters can be expressed in a prior density function $p_0(\phi)$. An example is to assign an uniform distribution to admissible values and zero density to physically impossible values of the parameters. This information can be combined with the the measurement data information contained in the likelihood function to form the posterior density function

$$p^*(\phi) = [\int L(\phi)p_0(\phi)\, d\phi]^{-1} L(\phi)p_0(\phi) \quad , \tag{2.11}$$

in which $p^*(\phi)$ stands for the density of the parameters ϕ after application of the information in the measured data. This follows from Bayes' theorem (Bard 1974). It can be seen as the joint probability density function of the likelihood function and the prior information. The first factor is needed for normalisation to obtain a proper probability density function.

This posterior density function, strictly not a probability density function because $p_0(\phi)$ is not, contains all elements of the probabilistic model, namely the deterministic model, the data, the probability distribution of the errors and the prior information on the parameters.

If it is theoretically and numerically possible to compute this posterior density function we still are not able to decide which parameter vector ϕ^* is the best estimate of the true vector $\tilde{\phi}$. The posterior density could have a shape like the one – dimensional function in figure 3. We are free to decide wether we take for instance: the mean

$$\bar{\phi} = \int_{-\infty}^{\infty} \phi\, p^*(\phi)\, d\phi \quad , \tag{2.12}$$

mode

$$\max p^*(\phi) \quad , \tag{2.13}$$

or median

$$\tilde{\phi} \mid \int_{-\infty}^{\tilde{\phi}} p^*(\phi)\, d\phi = \frac{1}{2} \tag{2.14}$$

of this distribution as the estimate ϕ^* .

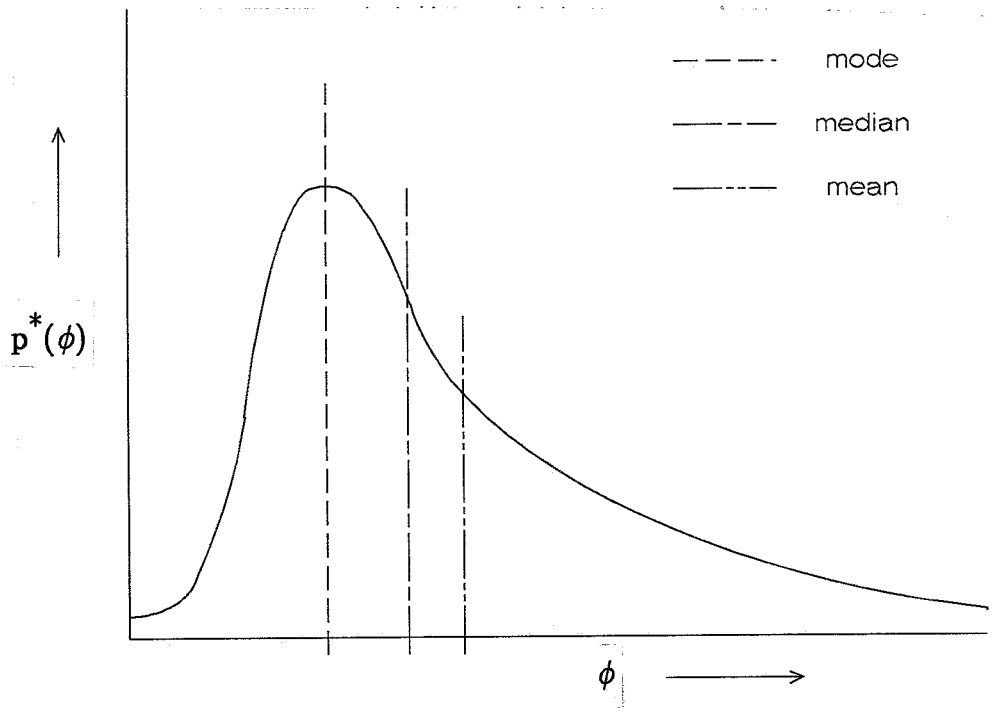

figure 3. Mean, mode and median of a distribution.

Decision theory can solve this dilemma, but has practical disadvantages and should only be used in real economic problems, as the definition of the cost or risk functions is quite arbitrary in an engineering approach.

3. MODEL ESTIMATION

As pointed out before, most commonly used procedures are point estimation methods. The functional relationship between the gathered data $\mathbf{Y}$ and the estimated parameter vector $\boldsymbol{\phi}^*$ is called the estimator and can be written as

$$\boldsymbol{\phi}^* = \mathbf{a}(\mathbf{Y}) \quad . \tag{3.1}$$

It will be clear that the deterministic estimator $\mathbf{a}$ will give raise to a random variable $\boldsymbol{\phi}^*$ because the input variable $\mathbf{Y}$ is of a stochastic nature. So the output variable will possess its own probability density distribution $p_a(\boldsymbol{\phi}^*)$, with its mean vector, covariance matrix and higher moments. At the model fitting level we only compute an estimate of the mean $\bar{\boldsymbol{\phi}}^*$ of the probability density distribution $p_a(\boldsymbol{\phi}^*)$. At the model estimation level we also try to compute an estimate of the covariance matrix $\mathbf{V}_{\boldsymbol{\phi}^*}$ of $p_h(\boldsymbol{\phi}^*)$.

In nonlinear estimation the variety in estimators in combination with their numerical implementations is beyound bound. For the assessment of one of these estimators some feavourable features can be defined, of which the two most important are mentioned here:

bias – The bias of an estimator is defined as $\mathbf{b} = \bar{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}$. A good estimator should be unbiased, that is the mean will equal the true values, or should have small bias. Due to the moderate sample size common estimators usually possess some bias, whenever this bias tends to zero (usually as 1/n) with infinite large sample size it is called consistent or asymptotically unbiased.

efficiency – The estimator should give the smallest possible covariance matrix, which is bounded below by the Rao – Cramer theorem. If it equals the theoretical bound the estimator is called efficient, if it approaches zero (usually linearly) with infinite large sample size it is called asymptotically efficient as is the case for most relevant estimators.

Only in the linear case these and other desired properties of the estimators can be derived or proved. In the nonlinear case these and other features of estimators can only be checked after that the estimator has been applied to the measured data. This can be done in two

different ways: the first is testing the estimator by repeating many test runs and comparison with exactly known systems which are hardly known, or by simulation of this process on digitally computers known as Monte Carlo techniques. By assuming a model and its parameters the "exact" measurements can be computed and compared with the "measured" values which can be obtained by means of the input of neighbouring parameters and added noise on the output variables. This will result in an estimate ϕ_i^* for each replication of the experiment (i = 1,2, . . . ,N), which is a sample of the distribution of the estimator (3.1). The mean, bias and the covariance matrix of the estimator can be computed from the well known formulas:

$$\bar{\phi}^* = \frac{1}{N} \sum_{i=1}^{N} \phi_i^* \quad , \tag{3.2}$$

$$\mathbf{b} = \bar{\phi}^* - \tilde{\phi} \quad , \tag{3.3}$$

$$\mathbf{V}_{\phi^*} = \frac{1}{N-1} \sum_{i=1}^{N} (\phi_i^* - \bar{\phi}^*)(\phi_i^* - \bar{\phi}^*)^T \quad . \tag{3.4}$$

The Monte Carlo method can be applied on fast computers and is very powerful and cost effective compared to real experiments.

The estimator itself usually consists of an objective function of the residuals which represents the lack of fit of the data to the chosen model. The estimator computes a minimum or a maximum of the defined objective function. In this stage all the numerical methods in the field of optimization can be applied (e.g. Gill, Murray & Wright 1981).

For the definition of estimators now we have to derive functions of the residuals taking into account various a priori known information. We will start the discussion with Bayesian estimation which uses the most known knowledge, however without risk functions, and gradually drop the need for prior information and end up with the most widely used method of least squares sums which can be applied directly to the deterministic model without any knowledge of the stochastic aspects.

Bayesian estimation

As discussed in the previous chapter, when using Bayesian estimation we should have knowledge of the model, the measurements, the error distribution and the prior density function of the parameters. We recall the posterior density function (2.11) in which all this information has been combined

$$p^*(\phi) = [\int L(\phi)p_o(\phi)\, d\phi]^{-1} L(\phi)p_o(\phi) \quad . \tag{3.5}$$

This expression only exists if the integral in the first factor can be calculated. The most commonly used estimate of this distribution is the mode or maximum value, because this value is the most simple to obtain from the posterior density function. So the objective function can be defined as equation (3.5). Most often it is advantageous to define the objective function by

$$\phi(\phi) = \ln(p^*(\phi)) \quad . \tag{3.6}$$

This can be done because the natural logarithm is a monotonic increasing function and so it will not change the optimal value. Application to probability distributions gives frequently a simpler expression. If the prior distribution does not vanish anywhere we have to maximize

$$\phi_b(\phi) = \ln(L(\phi)) + \ln(p_o(\phi)) \quad . \tag{3.7}$$

For the solution of this problem we have to apply a nonlinear optimization method for unconstrained problems. If it is vanishing somewhere, for instance due to an uniformly distributed prior density, we have to maximize a constrained problem, since not all parameter combinations will be feasible.

For normally distributed residuals and normally distributed prior knowledge concerning the parameters we can write with substitution of (2.10) in (3.7)

$$\phi_b(\phi) = -\frac{ns}{2}\ln(2\pi) - \frac{1}{2}\sum_{\mu=1}^{n} \ln \det V_\mu - \frac{1}{2}\sum_{\mu=1}^{n} e_\mu^T V_\mu^{-1} e_\mu -$$

$$-\frac{\ell}{2}\ln(2\pi)-\frac{1}{2}\ln(\prod_{\alpha=1}^{\ell}\sigma_\alpha^2)-\frac{1}{2}\sum_{\alpha=1}^{\ell}\frac{1}{\sigma_\alpha^2}(\phi_\alpha-\bar{\phi}_\alpha)^2 \quad , \tag{3.8}$$

in which the first terms of both lines and the second term on the second line can be ignored because they are constants. The second term in the first line should be retained because it contains the still unknown covariance matrices $\mathbf{V}_\mu$.

Maximum likelihood estimation

If we drop the prior knowledge of the parameter distribution and retain the information on the model, measurements and error distribution, the objective function can be defined by the logarithmic likelihood function

$$\phi_\ell(\boldsymbol{\phi}) = \ln(\mathrm{L}(\boldsymbol{\phi})) \quad , \tag{3.9}$$

which can be solved with unconstrained optimization procedures or constrained procedures if there exists simple bounds for the parameters. The maximum of this function (mode) is the most likely value of the parameters to occur in reality.

With normal distributed residuals the objective function that has to be maximized reads

$$\ln(\mathrm{L}) = -\frac{\mathrm{ns}}{2}\ln(2\pi)-\frac{1}{2}\sum_{\mu=1}^{\mathrm{n}}\ln\det\mathbf{V}_\mu-\frac{1}{2}\sum_{\mu=1}^{\mathrm{n}}\mathbf{e}_\mu^{\mathrm{T}}\mathbf{V}_\mu^{-1}\mathbf{e}_\mu \quad . \tag{3.10}$$

If we make a further assumption that the errors in all experiments have the same covariance matrix $\mathbf{V}$ formula (3.10) reduces to

$$\ln(\mathrm{L}) = -\frac{\mathrm{ns}}{2}\ln(2\pi)-\frac{\mathrm{n}}{2}\ln\det\mathbf{V}-\frac{1}{2}\sum_{\mu=1}^{\mathrm{n}}\mathbf{e}_\mu^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{e}_\mu =$$

$$-\frac{\mathrm{nr}}{2}\ln(2\pi)-\frac{\mathrm{n}}{2}\ln\det\mathbf{V}-\frac{1}{2}\mathrm{tr}(\mathbf{V}^{-1}\mathbf{M}(\boldsymbol{\theta})) \quad , \tag{3.11}$$

in which $\mathbf{M}(\boldsymbol{\theta})$ stands for the moment matrix of the residuals

$$\mathbf{M}(\boldsymbol{\theta}) = \sum_{\mu=1}^{n} \mathbf{e}_\mu \mathbf{e}_\mu^T \quad . \tag{3.12}$$

If we also assume that the errors within each experiment are independent, formula (3.11) will reduce to

$$\ln(L) = -\frac{ns}{2}\ln(2\pi) - \frac{n}{2}\sum_{\beta=1}^{s} \ln \sigma_\beta^2 - \frac{1}{2}\sum_{\beta=1}^{s} \frac{1}{\sigma_\beta^2}\mathbf{M}_{\beta\beta}(\boldsymbol{\theta}) \quad . \tag{3.13}$$

The last two optimization problems can be solved with derivative free methods, but also with stagewise maximization, see Bard 1974. The problem is divided into two stages, the first stage consists of the elimination of the distribution parameters and optimization to $\boldsymbol{\theta}$ and the second stage of the estimation of the eliminated distribution parameters.

Least squares estimation

In fact all types of least squares methods are simplifications of maximum likelihood estimators. In this case we drop the information on the error distribution and retain only the information of the model and the measurements. From an engineering point of view this is the same as applying the data directly to the deterministic model, which is the reason why it is used so frequently in engineering practice. So implicitly a normal distribution with known covariance matrices is assumed, because any knowledge on the error distribution is not available or hard to be obtained.

Now we can derive the likelihood functions for this last assumption. In formula (3.10) we can assume all $\mathbf{V}_\mu$ to be known, so the first and second term can be removed because they are constant. The objective function that has to be optimized reads

$$\ln(L) = C - \frac{1}{2}\sum_{\mu=1}^{n} \mathbf{e}_\mu^T \mathbf{V}_\mu^{-1} \mathbf{e}_\mu \quad . \tag{3.14}$$

Maximizing this function is equal to minimizing the well known weighted least squares sum

$$\phi(\theta) = +\frac{1}{2} \sum_{\mu=1}^{n} \mathbf{e}_\mu^T \mathbf{W}_\mu \mathbf{e}_\mu \quad , \tag{3.15}$$

in which $\mathbf{W}_\mu$ stands for the weight factor matrix which is inverse proportional to the covariance matrix of the residuals in equation (3.14). Weighted least squares estimates can also be found in literature as Markov estimates.

The last and best known unweighted least squares estimator, results from the assumption that all errors are independent with known variance σ^2 . Formula (3.13) reduces to

$$\ln(L) = C - \frac{1}{2\sigma^2} \sum_{\beta=1}^{s} \mathbf{M}_{\beta\beta}(\theta) = C - \frac{1}{2\sigma^2} \mathrm{tr}(\mathbf{M}(\theta)) \, . \tag{3.16}$$

Maximization of (3.16) is equivalent to minimization of the unweighted least squares formula

$$\phi(\theta) = +\frac{1}{2} \mathrm{tr}(\mathbf{M}(\theta)) = +\frac{1}{2} \sum_{\mu=1}^{n} \mathbf{e}_\mu^T \mathbf{e}_\mu \quad . \tag{3.17}$$

In least squares methods the weight factors generally are assigned by intuition or measurement equipment specifications what ever the real distribution parameters may be. Nevertheless these inverse variances or covariance matrices can be estimated from the data as shown in formulas (3.11) and (3.13). So it is preferable to use maximum likelihood methods if possible which are also able to estimate the unknown distribution parameters in stead of assignment by the user of some least squares method.

Sofar we have discussed estimation methods that only can be performed off – line after the complete run has been finished. In the field of control theory much work has been done on optimal estimation, in the least squares sense, of linear dynamical systems. These methods of estimation are called batch identification or smoothing, due to the off – line nature of the calculations. In real control problems one can not always use these estimation techniques because of the lack of time for computations or computer memory space limitations. The problem of tracking is defined by the on – line estimation of the state vector $\mathbf{s}(t_j)$ out of the measurements $\mathbf{y}_\mu$ $(\mu = 1,2, \ldots ,j)$, which can be seen as a form of filtering. In filtering one computes in a recursive way the current estimates out of a former solution by using the last available measurements. These methods are called sequential

estimation procedures and are usually based on the solution of some sort of Ricatti equation. For linear systems much work has been done by Kalman. His work is extensively discussed in literature and can be found in textbooks on control theory.

Covariance matrix of the estimates

We are able to estimate more characteristics than the mean value $\bar{\boldsymbol{\phi}}^*$ of the probability density function of the estimates (3.1), as in a point estimator discussed above. We can also compute an estimate of the covariance matrix of these parameters which reads

$$\mathbf{V}_{\boldsymbol{\phi}}{}^* = E\,[(\boldsymbol{\phi}^* - \bar{\boldsymbol{\phi}}^*)(\boldsymbol{\phi}^* - \bar{\boldsymbol{\phi}}^*)^{\mathrm{T}}] = E\,[\delta\boldsymbol{\phi}^*\ \delta\boldsymbol{\phi}^{*\mathrm{T}}] \quad . \tag{3.18}$$

If the estimator consists of one of the objective functions derived in the previous chapter, we can write down the following condition for the optimal value

$$\phi_{,\boldsymbol{\phi}}(\boldsymbol{\phi}^*,\mathbf{y}) = \mathbf{0} \quad , \tag{3.19}$$

where $\mathbf{y}$ is defined by (2.4). A variation $\delta\mathbf{y}$ in the measured variables $\mathbf{y}$ caused by two different realizations, will result in the desired variation $\delta\boldsymbol{\phi}^*$ in the optimal value of $\boldsymbol{\phi}^*$ in order to satisfy condition (3.19)

$$\phi_{,\boldsymbol{\phi}}(\boldsymbol{\phi}^* + \delta\boldsymbol{\phi}^*,\mathbf{y} + \delta\mathbf{y}) = \mathbf{0} \quad . \tag{3.20}$$

Expanding (3.20) in a Taylor series and retaining only the first order terms we can write after substracting (3.19)

$$\delta\boldsymbol{\phi}^* \approx \mathbf{H}^{*\,-1}\,\phi_{,\boldsymbol{\phi}\mathbf{y}}(\boldsymbol{\phi}^*,\mathbf{y})\ \delta\mathbf{y} \quad , \tag{3.21}$$

in which

$$\mathbf{H}^* = \phi_{,\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\phi}^*,\mathbf{y}) \quad . \tag{3.22}$$

This result can be substituted in the covariance matrix of the estimates (3.18)

$$\mathbf{V}_{\boldsymbol{\phi}}{}^* \approx E\,[\,\mathbf{H}^{*\,-1}\,\phi_{,\boldsymbol{\phi}\mathbf{y}}(\boldsymbol{\phi}^*,\mathbf{y})\ \delta\mathbf{y}\ \delta\mathbf{y}^{\mathrm{T}}\,\phi^{\mathrm{T}}_{,\boldsymbol{\phi}\mathbf{y}}(\boldsymbol{\phi}^*,\mathbf{y})\,\mathbf{H}^{*\,-1}\,] \quad , \tag{3.23}$$

that can be simplified to

$$\mathbf{V}_{\phi^*} \approx \mathbf{H}^{*\,-1}\, \phi,_{\phi\mathbf{y}}(\phi^*,\mathbf{y})\, \mathbf{V}_{\mathbf{y}}\, \phi^{\mathrm{T}},_{\phi\mathbf{y}}(\phi^*,\mathbf{y})\, \mathbf{H}^{*\,-1} \quad , \tag{3.24}$$

with $\mathbf{V}_{\mathbf{y}} = E\,[\delta\mathbf{y}\ \delta\mathbf{y}^{\mathrm{T}}]$, the covariance matrix of the measurements. If we can assume that the errors in different experiments are independent with covariance matrices $\mathbf{V}_{\mu}$ formula (3.24) can approximately be written as

$$\mathbf{V}_{\phi} \approx \mathbf{H}^{*\,-1} \sum_{\mu=1}^{n} [\, \phi,_{\phi\mathbf{y}}(\phi^*,\mathbf{y}_{\mu})\, \mathbf{V}_{\mu}\, \phi^{\mathrm{T}},_{\phi\mathbf{y}}(\phi^*,\mathbf{y}_{\mu})\,]\, \mathbf{H}^{*\,-1} \quad . \tag{3.25}$$

More specific results can be derived if further assumptions can be made (Bard 1974). Here only the most important formula will be discussed. If the objective function only depends on the moment matrix $\mathbf{M}(\boldsymbol{\theta})$, see equation (3.12), and Gauss approximation is employed, it can be shown that for a wide class of maximum likelihood estimates with normal distributions the following formula for the covariance matrix of the estimates approximately holds

$$\mathbf{V}_{\theta} \approx -\ln,_{\theta\theta}(\mathrm{L}(\boldsymbol{\theta}^*,\mathbf{y})) = \mathbf{H}^{*\,-1} \quad . \tag{3.26}$$

As this estimate is a sample from the distribution of $\mathbf{V}_{\theta}$ it is only of the right order of magnitude and should therefore only be used qualitatively to obtain an indicator of the goodness of fit of the data to the model. However it might be assumed that this estimate has quantitative value when convergence of the objective function has occurred to the most likely value of the parameters $\boldsymbol{\theta}^*$. For unweighted least squares estimators with Gauss approximation, see equation (3.16), this formula reduces to

$$\mathbf{V}_{\theta} \approx \mathbf{H}^{*\,-1} = \sigma^2\, (\mathbf{J}^{\mathrm{T}}\mathbf{J})^{-1} \quad . \tag{3.27}$$

Where $\mathbf{J}$ stands for the jacobian of the observed variables with respect to the parameters to be estimated as defined in equation (4.3) and σ^2 is the known variance or an estimate of the variance of the measured variables. The estimate can be computed according to

$$\sigma^2 = \frac{1}{n \times s - \ell} \sum_{\mu=1}^{n} \mathbf{e}_{\mu}^{\mathrm{T}}\, \mathbf{e}_{\mu} \quad . \tag{3.28}$$

During the optimization process no attention should be paid to this estimate of the covariance matrix. Only at the moment convergence has occurred, conclusions might be drawn.

Confidence regions

Theoretically, with the last information we are able to define confidence intervals in the one – dimensional or confidence regions in the multi – dimensional case for the estimates of the parameters $\boldsymbol{\theta}$. When convergence has been observed and when we can be sure of the normal distributed nature of the residuals, the confidence intervals belonging to the estimated parameters of unweighted least squares estimation can be derived. With application of the covariance matrix computed by equation (3.27) we simply can write the two – sided 95 % confidence interval for the exact parameter vector $\hat{\boldsymbol{\theta}}$ as

$$\theta_\alpha^* - 1.96\ \mathbf{V}_{\alpha\alpha}^{1/2} < \hat{\theta}_\alpha < \theta_\alpha^* + 1.96\ \mathbf{V}_{\alpha\alpha}^{1/2} \quad . \tag{3.29}$$

4. OPTIMIZATION

It is known in numerical optimization that it is advantageous to use the highest order of derivative information on the objective function as can reasonably be calculated. According to this statement we can give the following order of optimization methods in descending order of applicability and reliability.

If both the gradient and the Hessian can be computed against reasonable cost, *Newton methods with second order derivatives* can be used. When it is impossible or real expensive to compute the second order information we can use *Newton methods without second order information*, the Hessian is computed by finite differences of the gradient. This method is less reliable as we have to deal with the problem of balancing truncation and rounding errors. Another way to get some knowledge on the second order information is using one of the methods out of the large class of *quasi–Newton methods with first derivatives.* The idea of these methods is to build – up and store second order information during the iterations of some sort of descent procedure. Approximate second order information in one direction can be derived from the gradient vectors of two successive iterations and added to the last approximate Hessian. A well–known update is the Broyden – Fletcher – Goldfarb – Shanno (BFGS) update. Finite differences of the functional value can be applied to approximate the gradient in this method if the gradient is not available. This *quasi–Newton methods without first derivatives* suffer from the balancing of errors as mentioned with the Newton method. Especially when the problem is very large one could use the so – called *conjugate gradient method*, respectively again with or without first derivatives. These methods are specifically suited for large and/or sparse problems because they approximate the Hessian implicitly. At every iteration the direction of search is computed directly without storage of the Hessian or its inverse. The last class of methods that should not be used preferably, despite of their simplicity, are *direct search methods* (e.g. the simplex or polytope method). These methods are designed for optimization of non – smooth functions and make no use of any derivative information. Due to lack of derivative information and the exponential growth of function evaluations with higher dimensions they appear to be the least efficient algorithms.

In parameter estimation we are usually faced with the optimization of objective functions of which the first and sometimes the second derivatives can be derived and computed. So

we are able to use Newton–like methods. Especially when we are dealing with some sort of least squares estimation we have the opportunity to use dedicated optimization algorithms. Suppose that the objective function is the unweighted nonlinear least squares sum (3.17)

$$\phi(\boldsymbol{\theta}) = +\frac{1}{2}\,\mathrm{tr}(\mathbf{M}(\boldsymbol{\theta})) = +\frac{1}{2}\sum_{\mu=1}^{n} \mathbf{e}_{\mu}^{T}\,\mathbf{e}_{\mu} = +\frac{1}{2}\,\|\,\mathbf{y}-\hat{\mathbf{y}}\,\|_2^2 \quad . \tag{4.1}$$

From a statistical point of view the $2-$norm followed from the assumption of normally distributed errors. From a mathematical point of view this $2-$norm has some feavourable features compared to other norms (e.g $1-$norm, sum of the elements or $\infty-$norm, the largest element), that becomes clear in the derivation of the gradient and the Hessian of the objective function

$$\phi(\boldsymbol{\theta}+\delta\boldsymbol{\theta}) \approx \phi(\boldsymbol{\theta}) + \delta\boldsymbol{\theta}^{T}\mathbf{q}(\boldsymbol{\theta}) + \frac{1}{2}\,\delta\boldsymbol{\theta}^{T}\mathbf{H}(\boldsymbol{\theta})\delta\boldsymbol{\theta} \quad , \tag{4.2}$$

with

$$\mathbf{q}(\boldsymbol{\theta}) = \sum_{\mu=1}^{n} \mathbf{e}_{\mu,\boldsymbol{\theta}}^{T}\,\mathbf{e}_{\mu} = -\hat{\mathbf{y}}_{,\boldsymbol{\theta}}^{T}(\mathbf{y}-\hat{\mathbf{y}}) = -\mathbf{J}^{T}(\boldsymbol{\theta})\,\mathbf{e} \quad , \tag{4.3}$$

$$\mathbf{H}(\boldsymbol{\theta}) = \sum_{\mu=1}^{n} \mathbf{e}_{\mu,\boldsymbol{\theta}}^{T}\,\mathbf{e}_{\mu,\boldsymbol{\theta}} + \sum_{\mu=1}^{n}\sum_{\alpha=1}^{\ell} \mathbf{e}_{\mu,\boldsymbol{\theta}\theta_{\alpha}}^{T}\,\mathbf{e}_{\mu} =$$

$$= \mathbf{J}^{T}(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta}) + \mathbf{Q}(\boldsymbol{\theta}) \quad . \tag{4.4}$$

The solution of the normal equations for this problem gives us the Newton direction of search $\mathbf{p}_n$ for iteration i

$$(\mathbf{J}^{T}(\boldsymbol{\theta})\mathbf{J}(\boldsymbol{\theta}) + \mathbf{Q}(\boldsymbol{\theta}))_i\,\mathbf{p}_{ni} = \mathbf{J}_i^{T}(\boldsymbol{\theta})\,\mathbf{e}_i \quad . \tag{4.5}$$

This direction is the optimal solution of the complete quadratic local model of the nonlinear objective function in the neighbourhood of $\boldsymbol{\theta}_{i-1}$.

In general we may now assume that with nonlinear least squares problems and good approximate parameter vectors $\boldsymbol{\theta}$ the residuals $\mathbf{e}_{\mu}$ will be small, at least at the optimal

value θ^* . So we can ignore the second order information $Q(\theta)$ to the squared first order information in the Hessian (4.4). The solution of the normal equations for the remaining problem gives us the Gauss – Newton direction of search p_{gn} for iteration i

$$(J^T(\theta)J(\theta))_i \, p_{gni} = J_i^T(\theta) \, e_i \quad . \tag{4.6}$$

The decomposition of this set of linear equations has the disadvantage of squaring the condition number of $J(\theta)$ which can be fairly high anyway. Equation (4.6) can be solved conveniently by employment of the pseudo – inverse matrix $J(\theta)^+$ which can be computed by singular value decomposition. The condition number of the pseudo – inverse matrix will be of the same order of magnitude as that of matrix $J(\theta)$. The solution reads

$$p_{gni} = (J^T(\theta)J(\theta))_i^{-1} \, J_i^T(\theta) \, e_i = J(\theta)_i^+ \, e_i \quad . \tag{4.7}$$

When a direction of search has been found we can proceed with an optimization in that direction. This is called a linear search and the total method will be called a step – length – based method. The resulting step length will usually differ from unity as in the pure Newton or Gauss – Newton methods, because the real model will not exactly be represented by the local quadratic model. Another frequently used simplification is to use under – relaxation in the first iterates to obtain convergence.

A very successful method is the Levenberg – Marquardt method. It computes a direction of search p_{lmi} from the modified normal equations

$$(J^T(\theta)J(\theta) + \lambda_i \, I)_i \, p_{lmi} = J_i^T(\theta) \, e_i \quad , \tag{4.8}$$

in which λ_i is a non – negative scalar. The step length always equals unity in this method. It can be proved that the solution of the unconstrained system (4.8) is equal to the solution of the constrained system

$$\text{minimize} \quad \frac{1}{2} \| - J_i^T(\theta) \, p_{lmi} + e_i \|_2^2 \quad , \tag{4.9}$$

$$\text{subject to} \quad \| \, p_{lmi} \, \| \leq \Delta_i$$

where Δ_i is a scalar related to λ_i . This is an example of a trust – region type of

optimization. The scalar Δ_i , or equivalently λ_i , limits the value of the step length in an admissable region around the current solution $\boldsymbol{\theta}_i$ and must be choosen in such a way to ensure an optimal descent direction and positive definite Hessian matrix. There exists several algorithms to adjust the value of the scalar during the optimization process. Two extreme values of λ_i of this method, namely $\lambda_i \to 0$ and $\lambda_i \to \infty$ result respectively in the Gauss – Newton and steepest descent directions (opposite to the gradient).

The methods discussed sofar all assume neglegible residuals during the optimization process. When large residuals are encountered, we do need better second order information than the square of the jacobian of the predicted values alone. If it is not possible or to expensive to compute the total Hessian according to (4.4) we could employ quasi Gauss – Newton or conjugate – gradient algorithms both with analytical first order derivatives. If the objective function is that complicated that we are not able to derive the analytical derivative we have to employ quasi Gauss – Newton or conjugate – gradient algorithms with finite difference first order derivatives approximations.

The final method that will be mentioned here is a modified Gauss – Newton method that is capable of handling large residuals and rank – deficient problems. The method makes use of singular value decomposition of the Jacobian $\mathbf{J}(\boldsymbol{\theta})$

$$\mathbf{J}(\boldsymbol{\theta}) = \mathbf{U} \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix} \mathbf{V}^T \quad , \tag{4.10}$$

in which $\mathbf{S}$ represents a diagonal matrix containing the singular values $(\sigma_1, \sigma_2, \ldots, \sigma_\ell)$ $\sigma_i \geq \sigma_{i+1}$ and $\mathbf{U}$ a (n×m)×(n×m) orthonormal matrix and $\mathbf{V}$ a $\ell \times \ell$ orthonormal matrix. Substitution in the normal equations (4.5) gives

$$(\mathbf{V} [\mathbf{S}^T \; 0] \; \mathbf{U}^T \; \mathbf{U} \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix} \mathbf{V}^T + \mathbf{Q})_i \; \mathbf{p}_{ni} = (\mathbf{V} [\mathbf{S}^T \; 0] \; \mathbf{U}^T)_i \; \mathbf{e}_i \quad , \tag{4.11}$$

which holds for every iteration i . The index will be omitted in the following derivation for simplicity. With use of the orthonormality of $\mathbf{U}$ and $\mathbf{V}$ we can write

$$(\mathbf{S}^2 + \mathbf{V}^T \mathbf{Q} \mathbf{V}) \; \mathbf{V}^T \mathbf{p}_n = \mathbf{S} \, \mathbf{e}_\ell \quad , \tag{4.12}$$

in which $\mathbf{e}_\ell$ stands for the partition of the first ℓ elements of $\mathbf{U}^T \mathbf{e}$. So the Gauss – Newton direction can be computed by means of singular value decomposition, neglecting

the second order terms $\mathbf{V}^T\mathbf{Q}\mathbf{V}$ as mentioned in the discussion on equation (4.7)

$$\mathbf{p}_{gn} = \mathbf{V}\mathbf{S}^{-1}\,\mathbf{e}_\ell = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^T\,\mathbf{e} \quad . \tag{4.13}$$

In the case that second order information can not be neglected due to a rank – deficient jacobian or the magnitude of the second order information itself, we can devide the singular values into two partitions. The first partition contains the p1 dominant singular values, the second the $p2 = \ell - p1$ zero or small singular values. The value of p1 is also called the grade of the singular value decomposition. All other equations can be partitioned according to the defined sets. The direction of search can be written as a linear combination of the first p1 and second p2 columns of $\mathbf{V}$, $\mathbf{V}_1$ and $\mathbf{V}_2$ respectively

$$\mathbf{p}_n = \mathbf{V}_1\,\mathbf{p}_1 + \mathbf{V}_2\,\mathbf{p}_2 \quad . \tag{4.14}$$

Making use of the orthonormality of $\mathbf{V}$ we can partition equation (4.12) as follows

$$\left[\begin{bmatrix}\mathbf{S}_1^2 & 0\\ 0 & \mathbf{S}_2^2\end{bmatrix} + \begin{bmatrix}\mathbf{V}_1^T\\ \mathbf{V}_2^T\end{bmatrix}\mathbf{Q}\,[\mathbf{V}_1\;\mathbf{V}_2]\right]\begin{bmatrix}\mathbf{p}_1\\ \mathbf{p}_2\end{bmatrix} = \begin{bmatrix}\mathbf{S}_1\mathbf{e}_\ell\\ \mathbf{S}_2\mathbf{e}_\ell\end{bmatrix} \quad . \tag{4.15}$$

The equations of the first partition of (4.15) are well conditioned, so they can be solved without second order information. In the second partition the second order information will be necessary to compute a good estimate for the direction of search in this subset of the parameter space. The first part of the direction of search $\mathbf{V}_1\mathbf{p}_1$ in the subspace $\mathbf{V}_1$ can be approximated with use of $\mathbf{p}_{1gn}$, that is the solution of the Gauss – Newton formula of (4.15 – a)

$$\mathbf{p}_{1gn} = \mathbf{S}_1^{-1}\,\mathbf{e}_\ell \quad . \tag{4.16}$$

In the subspace $\mathbf{V}_2$ we can compute the second partition of the direction of search from equation (4.15 – b), with substitution from $\mathbf{p}_{1gn}$ for $\mathbf{p}_1$.

$$(\mathbf{S}_2^2 + \mathbf{V}_2^T\,\mathbf{Q}\,\mathbf{V}_2)\,\mathbf{p}_2 = \mathbf{S}_2\,\mathbf{e}_\ell - \mathbf{V}_2^T\,\mathbf{Q}\,\mathbf{V}_1\,\mathbf{p}_{1gn} \quad . \tag{4.17}$$

The necessary second order information $\mathbf{Q}$ need not be computed explicitly, as we can gather this information by finite difference techniques to the gradient along the columns of

$\mathbf{V}_2$ to approximate $\mathbf{Q}\,\mathbf{V}_2$ or by application of quasi Gauss – Newton techniques. The total solution of this modified Gauss – Newton can now be written as

$$\mathbf{p}_{mgn} = \mathbf{V}_1\,\mathbf{p}_{1gn} + \mathbf{V}_2\,\mathbf{p}_2 \quad . \tag{4.18}$$

This method can be seen as a compromise between the Newton and Gauss – Newton techniques as the extreme values $p1 = 0$ and $p1 = \ell$ respectively represents both methods. The idea of this method is not to use the grade of the singular value decomposition directly for the value of $p1$, but to set the value to ℓ as long as sufficient progress is being made and to update the value otherwise during the optimization.

Discussion on methods for linear and nonlinear constrained optimization would go beyound the scope of this report and can be found in Gill, Murray & Wright (1981), in chapters 5 and 6 respectively. Most of these methods are merely based on iteratively solving the projected quadratic model on the feasible or the constrained directions.

5. DYNAMICAL SYSTEMS

In dynamical system estimation it is evident that by using laws of physics differential equations occur in the mathematical description of the process. This consideration will lead to another derivation of the residuals needed in the likelihood functions. In addition to this, the exact solution of these equations according to an assumed parameter vector can be obtained by two quite different solution techniques in dynamics, which will have its effects on the computation of the gradient of the objective function. The first estimation method for dynamical systems to be discussed in this chapter is the method using the discrete time solution computed by time integration procedures, the second one is a method employing the discrete time solution of a periodic solver.

5.1 Time integration method

In general the deterministic dynamical model can be written as

$$\mathbf{g}(\mathbf{u},\dot{\mathbf{u}},\ddot{\mathbf{u}},\mathbf{x},t,\boldsymbol{\theta}) = 0 \quad , \tag{5.1}$$

in which $\mathbf{u}$ stands for the displacements, $\mathbf{x}$ for the independent variables and t for time. The vector $\boldsymbol{\theta}$ denotes the parameter vector. According to the order of the system a sufficient number of initial conditions must be supplied or estimated too. Nowadays in structural dynamical systems the state space description is widely used

$$\dot{\mathbf{s}} = \mathbf{h}(\mathbf{s},\mathbf{x},t,\boldsymbol{\theta}) \quad . \tag{5.2}$$

The column $\mathbf{s}$ denotes the m – dimensional state vector, $\mathbf{x}$ represents the r – dimensional independent variable vector (e.g. forcing functions), t stands for time and $\boldsymbol{\theta}$ is the ℓ – dimensional to be estimated parameter vector. Of course the necessary initial conditions must be added

$$\mathbf{s}(0) = \mathbf{s}_0(\mathbf{x},\boldsymbol{\theta}) \quad , \tag{5.3}$$

which might be a function of the independent variables (known conditions) or the unknown

parameters themselves. The last set of equations that completes the description, are the output equations that relate the observed variables to the complete set of other variables

$$\hat{\mathbf{y}} = \mathbf{f}(\mathbf{s},\mathbf{x},t,\boldsymbol{\theta}) \quad . \tag{5.4}$$

The vector $\hat{\mathbf{y}}$ again denotes the exact values of the observed variables. Note that the parameters can occur in all equations: as parameters in the differential equations, as unknown initial conditions and as parameters in the observation model.

Solution of the differential equations (5.2) together with the initial conditions (5.3) gives the state vector $\mathbf{s}(t)$ which can be substituted in the observation model (5.4). These equations will simplify to the explicit or implicit relations

$$\hat{\mathbf{y}} = \mathbf{f}(\mathbf{x},t,\boldsymbol{\theta}) \quad . \tag{5.5}$$

On the other hand we have the experiments. The measured variables can be written as functions of the independent variables $\mathbf{x}$ and time t

$$\mathbf{y} = \mathbf{f}(\mathbf{x},t) \quad . \tag{5.6}$$

A group of measurements that belong to the same set of independent variables $\mathbf{x}$ and initial conditions, thus differ only in time t, will be called a run. Each run gives raise to a seperate estimation problem. In this case, the index μ ($\mu = 1,2, \ldots ,n$) will be used to label the observed variables at time t_μ in a run. The gathered data of a run can again be stored in vectors $\mathbf{y}_\mu$, in a matrix $\mathbf{Y}$ or in a vector $\mathbf{y}$.

Now we can write the residuals for this problem as the difference between the the measured and predicted values according to a feasible parameter vector $\boldsymbol{\theta}$ as

$$\mathbf{e}_\mu = \mathbf{y}_\mu - \hat{\mathbf{y}}_\mu \quad , \tag{5.7}$$

or for a complete run in matrix notation as $\mathbf{E}$ or $\mathbf{e}$, with which most objective functions, such as likelihood functions and least squares sums, can be derived.

In solving optimization problems it is recommended to use all the information which could be obtained in a reasonable way (Gill, Murray, Wright 1981; § 8.1). In optimizing

procedures it is often advantageous to have derivative information on the objective function with subject to the unknown parameters. The gradient of an objective function can be computed in two different ways:

finite differences – The variation of the objective function $\Delta\phi$ due to a variation in the parameters $\Delta\boldsymbol{\theta}$ could be computed and the ratio of both, $\Delta\phi/\Delta\boldsymbol{\theta}$, can be used for an approximation of the exact gradient $\phi_{,\boldsymbol{\theta}}$. In this solution we have to deal with the numerical problem of balancing the truncation and the rounding errors, which respectively increase and decrease the total error with increasing $\Delta\boldsymbol{\theta}$. This method is computer time consuming because mutiple precision or high order integration will be acquired to reach a satisfying solution to the problem.

Sensitivity equations – Linearization of the nonlinear differential equations about the trajectory belonging to an estimate parameter vector gives a variant linear set of differential equations which can be integrated along with the state equations. In this case there are no problems with accuracy as the two sets of equations can be integrated with the same method and step sizes. In contrast with finite differences, when using this method the gradient can be computed with the same accuracy as the objective function itself.

In order to calculate the gradient $\phi_{,\boldsymbol{\theta}}$ we have to know all the dependencies of the objective function to the parameter vector:

ϕ depends on $\mathbf{e}_\mu = \mathbf{y}_\mu - \hat{\mathbf{y}}_\mu$ and maybe on $\boldsymbol{\theta}$ directly if prior information has been added to the objective function.

$\hat{\mathbf{y}}_\mu$ depends on $\mathbf{s}_\mu$ and $\boldsymbol{\theta}$ by the output relations (5.4).

$\mathbf{s}_\mu$ depends on $\mathbf{s}_0$ and on $\boldsymbol{\theta}$ by integration of the state equations (5.2) and the initial conditions (5.3).

$\mathbf{s}_0$ might depend on $\boldsymbol{\theta}$ trough equation (5.3).

Now we can derive the gradient of the objective function (without prior information)

$$q = \phi_{,\theta} = \sum_{\mu=1}^{n} (\phi_{,e_\mu}^T e_{\mu,\theta})^T = - \sum_{\mu=1}^{n} \hat{y}_{\mu,\theta}^T \phi_{,e_\mu} =$$

$$- \sum_{\mu=1}^{n} (f_{\mu,\theta} + f_{\mu,s_\mu} s_{\mu,\theta})^T \phi_{,e_\mu} \quad . \tag{5.8}$$

If we apply prior information to the objective function we have to add $\bullet \ln,_\theta(p_0)$ to the gradient, the positive sign for a proper log – likelihood function that has to maximized and the negative sign for a least squares sum that has to be minimized.

The partial derivatives $\partial\phi/\partial e_\mu$ and $\partial f_\mu/\partial\theta$, $\partial f_\mu/\partial s_\mu$ can in most cases be derived analytically from the objective function and output equations respectively. The remaining derivatives $\partial s_\mu/\partial\theta$ can be calculated from the socalled sensitivity equations that can be derived in the following way. Recall the state equations (5.2)

$$\dot{s} = h(s,x,t,\theta) \quad .$$

Differentiation with respect to the parameters gives

$$(\dot{s}),_\theta = h,_\theta + h,_s s,_\theta \quad . \tag{5.9}$$

Changing the sequence of differentiations on the left hand side of the equation reads

$$(\dot{s},_\theta) = h,_s s,_\theta + h,_\theta \quad , \tag{5.10}$$

which are called the sensitivity equations. The coefficients $h,_s$ and $h,_\theta$ can with some effort be derived from the state equations. With the corresponding initial conditions $s,_\theta(0)$ we can integrate this set of equations simultaneously with the state equations to compute $\partial s_\mu/\partial\theta$ to complete the information for the calculation of the objective function gradient. This method gives better values for the gradient as finite differencing but it should be clear that it requires more effort to form the sensitivity equations analytically. This problem might be solved by a suitable symbolic computer program.

When applying this method in one of the many optimization procedures available one should keep in mind that whenever variable step methods are used only the state equations may influence the step length. In most optimization methods there are two types of

iterates, at a main iterate the functional value and the gradient are calculated, in the linear search however only the functional value. At every point in parameter space the functional value should be unique to ensure deterministic behaviour of the objective function to obtain normal convergence. The disadvantage of the less optimal gradient value should be taken for granted.

In the literature a variant of this method can be found (Batill 1988, Goodman 1966). These authors replace the unknown parameter vector $\boldsymbol{\theta}$ by an additional new state vector $\mathbf{s}_{\boldsymbol{\theta}}$ subject to the additional state equations

$$\dot{\mathbf{s}}_{\boldsymbol{\theta}} = \mathbf{0} \quad , \tag{5.11}$$

with

$$\mathbf{s}_{\boldsymbol{\theta}}(0) = \boldsymbol{\theta} \quad . \tag{5.12}$$

The description of the problem might seem to simplify the problem by decreasing the number of derivatives that we have to derive, but in fact it is unfavourable for the total cpu – time. This can be understood if one realizes that the number of sensitivity equations to be integrated have been enlarged by the additional state variables. Integrating these additional equations will give no extra information as the value of the additional $\partial \mathbf{s}_{\boldsymbol{\theta}\mu} / \partial \boldsymbol{\theta}$ will not change and keep its identity matrix value during the integration process.

5.2 Periodic solution method.

The second method mentioned in the introduction of this chapter makes use of the periodic solutions of equation (5.1)

$$\mathbf{g}(\mathbf{u},\dot{\mathbf{u}},\ddot{\mathbf{u}},\mathbf{x},t,\boldsymbol{\theta}) = \mathbf{0} \quad ,$$

in which $\mathbf{x}$, the independent variables, contains the periodic excitation forces of the equations of motion of the system. In the sequel a brief outline of the method is presented. For a more detailed description see Crooijmans (1987), Hoogstraten (1987). For periodic solutions the problem can be rescaled to a cyclic time variable

$$\tau = t\, f_e \quad , \quad \tau \in [0,1> \quad , \tag{5.13}$$

in which f_e denotes the excitation frequency. According to this definition the dependent variables (positions) and its derivatives can be written as

$$\mathbf{u}(t) = \mathbf{z}(\tau) \quad , \tag{5.14}$$

$$\dot{\mathbf{u}}(t) = f_e\, \mathbf{z}'(\tau) \quad ,$$

$$\ddot{\mathbf{u}}(t) = f_e^2\, \mathbf{z}''(\tau) \quad .$$

These definitions can be substituted in the equations of motion (5.1), which read

$$\mathbf{g}(\mathbf{z}(\tau),\mathbf{z}'(\tau),\mathbf{z}''(\tau),\mathbf{x},\tau,\boldsymbol{\theta}) = \mathbf{0} \quad , \tag{5.15}$$

To obtain a discrete periodic solution, the solution $\mathbf{z}(\tau)$ can be discretized in time in n steps and the resulting variables can be stored in the column

$$\mathbf{z} = [\mathbf{z}^T(\tau_1),\mathbf{z}^T(\tau_2), \ldots ,\mathbf{z}^T(\tau_n)]^T \quad . \tag{5.16}$$

By employment of finite difference schemes the derivatives of the solution $\mathbf{z}(\tau)$ can be written as functions of the vector of variables $\mathbf{z}$, for instance with a second order scheme

$$\mathbf{z}'(\tau_i) = \frac{1}{2d}\left(-\mathbf{z}(\tau_{i-1}) + \mathbf{z}(\tau_{i+1})\right) \quad , \tag{5.17}$$

$$\mathbf{z}''(\tau_i) = \frac{1}{d^2}\left(\mathbf{z}(\tau_{i-1}) - 2\mathbf{z}(\tau_i) + \mathbf{z}(\tau_{i+1})\right) \quad ,$$

$$\text{in which } d = 1/n \quad .$$

Substitution of these expressions in the equations of motion (5.15) gives for each time step τ_μ a set of equations

$$\mathbf{g}_\mu(\mathbf{z},\mathbf{x},\tau_\mu,\boldsymbol{\theta}) = \mathbf{0} \quad , \tag{5.18}$$

or for all time steps a set of nonlinear algebraic equations

$$\mathbf{g}(\mathbf{z},\mathbf{x},\tau,\boldsymbol{\theta}) = [\mathbf{g}_1^T,\mathbf{g}_2^T,\ldots,\mathbf{g}_n^T]^T = \mathbf{0} \quad . \tag{5.19}$$

The last set of equations can be solved for $\mathbf{z}$ by a Newton – Raphson procedure.

This deterministic model again can be completed with a set of output equations that relate the observed variables to the complete set of other variables in this system

$$\hat{\mathbf{y}} = \mathbf{f}(\mathbf{z},\mathbf{x},t,\boldsymbol{\theta}) \quad . \tag{5.20}$$

The measurements themselves can be written as

$$\mathbf{y} = \mathbf{f}(\mathbf{x},t) \quad . \tag{5.21}$$

For the same reasons as in the previous section it is advantageous to derive the first order derivative information of the objective function of the residuals to be minimized. For an objective function without prior information we can write this particular gradient as

$$\mathbf{q} = \phi_{,\boldsymbol{\theta}} = \sum_{\mu=1}^{n} (\phi_{,\mathbf{e}_\mu}^T \mathbf{e}_{\mu,\boldsymbol{\theta}})^T = -\sum_{\mu=1}^{n} \hat{\mathbf{y}}_{\mu,\boldsymbol{\theta}}^T \phi_{,\mathbf{e}_\mu} =$$

$$-\sum_{\mu=1}^{n} (\mathbf{f}_{\mu,\boldsymbol{\theta}} + \mathbf{f}_{\mu,\mathbf{z}}\, \mathbf{z}_{,\boldsymbol{\theta}})^T \phi_{,\mathbf{e}_\mu} \quad . \tag{5.22}$$

The partial derivatives $\partial\phi/\partial\mathbf{e}_\mu$ and $\partial\mathbf{f}_u/\partial\boldsymbol{\theta}$, $\partial\mathbf{f}_\mu/\partial\mathbf{z}_\mu$ can be derived analytically from the objective function and output equations respectively. The only remaining unknowns $\partial\mathbf{z}_\mu/\partial\boldsymbol{\theta}$, can be seen as partitions of the matrix

$$\mathbf{z}_{,\boldsymbol{\theta}} = [\mathbf{z}_{,\boldsymbol{\theta}}^T(\tau_1);\mathbf{z}_{,\boldsymbol{\theta}}^T(\tau_2);\ldots;\mathbf{z}_{,\boldsymbol{\theta}}^T(\tau_n)]^T \quad , \tag{5.23}$$

which can be computed from an analogon of the sensitivity equations (5.10) for this periodic solution method. The sensitivity equations can be derived by differentiation of the nonlinear algebraic equations (5.19) with respect to the parameter vector $\boldsymbol{\theta}$

$$\mathbf{g}_{,\boldsymbol{\theta}}(\mathbf{z},\mathbf{x},\tau,\boldsymbol{\theta}) = \mathbf{g}_{,\boldsymbol{\theta}} + \mathbf{g}_{,\mathbf{z}}\, \mathbf{z}_{,\boldsymbol{\theta}} = \mathbf{0} \tag{5.24}$$

$$g_{,z}\, z_{,\theta} = -g_{,\theta} \quad .$$

The derivatives $\partial g / \partial z$ can be extracted from the periodic solver as they were needed in the Newton –Raphson process. The partial derivatives $\partial g / \partial \theta$ have to be derived analytically or by symbolic computation from equations (5.19).

With the last information we are able to define a periodic parameter estimation algorithm. The equations to be solved consist of an optimizable objective function resulting from one of the chosen estimation techniques of chapter 3. The functional values of this objective function can be computed by application of the solution of the set of nonlinear algebraic equations (5.19) and the measured data. The gradient of the objective function can even be obtained with some computational effort from equation (5.22). Both the functional value and the gradient can be input to one of the optimization methods of chapter 4 or even more sophisticated methods to approximate the optimal parameter vector for a dynamical system.

6. ESTIMATION OF AN OLEO

A literature search on estimation or identification of landing gear was not very successful. Just a single article by Batill & Bacarro (1988) could be traced, dealing with this particular subject. The reported method of estimation is a response curve method. The predicted response, integrated according to an estimated parameter vector and initial conditions, is compared with a measured response or other output signals. An unweighted least squares estimator with Gauss approximation is employed in order to compute the next update of the parameter vector and initial conditions. As stated in the introduction several problems were encountered such as: divergence by ill – conditioning, sensibility to scaling of the variables, numerical overflow by inadmissable parameter combinations or discontinuities in friction functions.

The first problem can have two different origins, namely inherent ill – conditioned differential equations or ill – conditioning of the problem by using nonoptimal measurement signals. A very simple example of an ill – conditioned differential equation is

$$\ddot{u} + \theta_1 u + \theta_2 u = \ddot{u} + (\theta_1 + \theta_2) u = 0 \quad , \tag{6.1}$$

in which both θ_1 and θ_2 , of course, can not be estimated independently. However other less easy detectable forms of this type off ill – conditioning can occur. One could think of combinations of different types of friction in nonlinear problems.

The ill – conditioning in the article of Batill & Bacarro is caused by nonoptimal sampled signals. It is a well known fact that the Shannon theorem should be satisfied, it orders the sample frequency to be at least the double of the highest frequency in the original signal. Another condition that must be satisfied is that the sampled signal must contain sufficient information of the system under consideration. The first condition is met in the article by a fixed sample frequency of 100 Hz , but the second is not.

Looking at the results of Batill's report on the identification of a linear oscillator with viscous damping, defined by the following differential equation

$$\ddot{u} + \theta_1 \dot{u} + \theta_2 u = 0 \quad , \tag{6.2}$$

we can expect the response of this oscillator, according to the following parameters, to be like figure 4.

Parameter	Value
u(0)	1.00 inches
$\dot{u}(0)$	0.00 inch/s
θ_1	0.25 s^{-1}
θ_2	7.50 s^{-2}

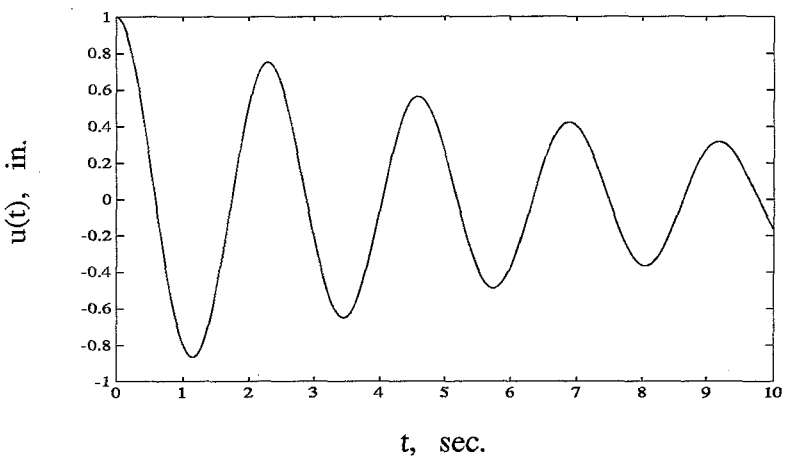

figure 4. Typical free response of a linear oscillator with viscous damping.

The noise free displacement response is simulated using a fourth order Runge – Kutta integration scheme. The discretisation time step is 0.01 seconds, so that the sample frequency is 100 Hz . Batill used a fixed sample time step in his identifications and studied the results of using 4, 50, 100, 250 and 500 measurement points. He concluded that when he was using fewer measurements, the estimation procedure was failing due to ill – conditioning. This is only true when the measurement points are constrained to a fixed sample frequency. Figures 5 to 7 show the identification iterations of a measurement signal of 5 points with corresponding parameter vectors and initial conditions. In figure 5 the end time is only 0.04 seconds so that we have the first 5 samples at a sample rate of 100 Hz of the damped sine wave of figure 4. The solid line is the exact solution with 5% added uniform noise, the dashed line is the response corresponding to the initial guess of the parameters which are in error by +5% and the dotted line is the computed response belonging to the estimated converged parameter vector. In this case the converged parameter vector is indeed worthless and caused by ill – conditioning of the jacobian of the normal equations. However solutions of 5 measurement points with lower sample rates of 2 Hz in figure 6 and 1 Hz in figure 7 show that the identification can be improved if the time signal contains relevant information of the response of the system. In figures 5 to 7 the simulated time intervals contain 0.1, 1 and 2 cycles respectively.

measurement time : 0.04 [s]
number of measurement points : 5
noise level : 5 %

exact parameters		estimated parameters		1.0e+003 *	
1	0.0011	0.0010	0.0010	0.0010	0.0010
0	0	0.0052	0.0052	0.0052	0.0052
0.25	0.0003	4.5175	4.5175	4.5175	4.5175
7.5	0.0079	0.6588	0.6588	0.6588	0.6588

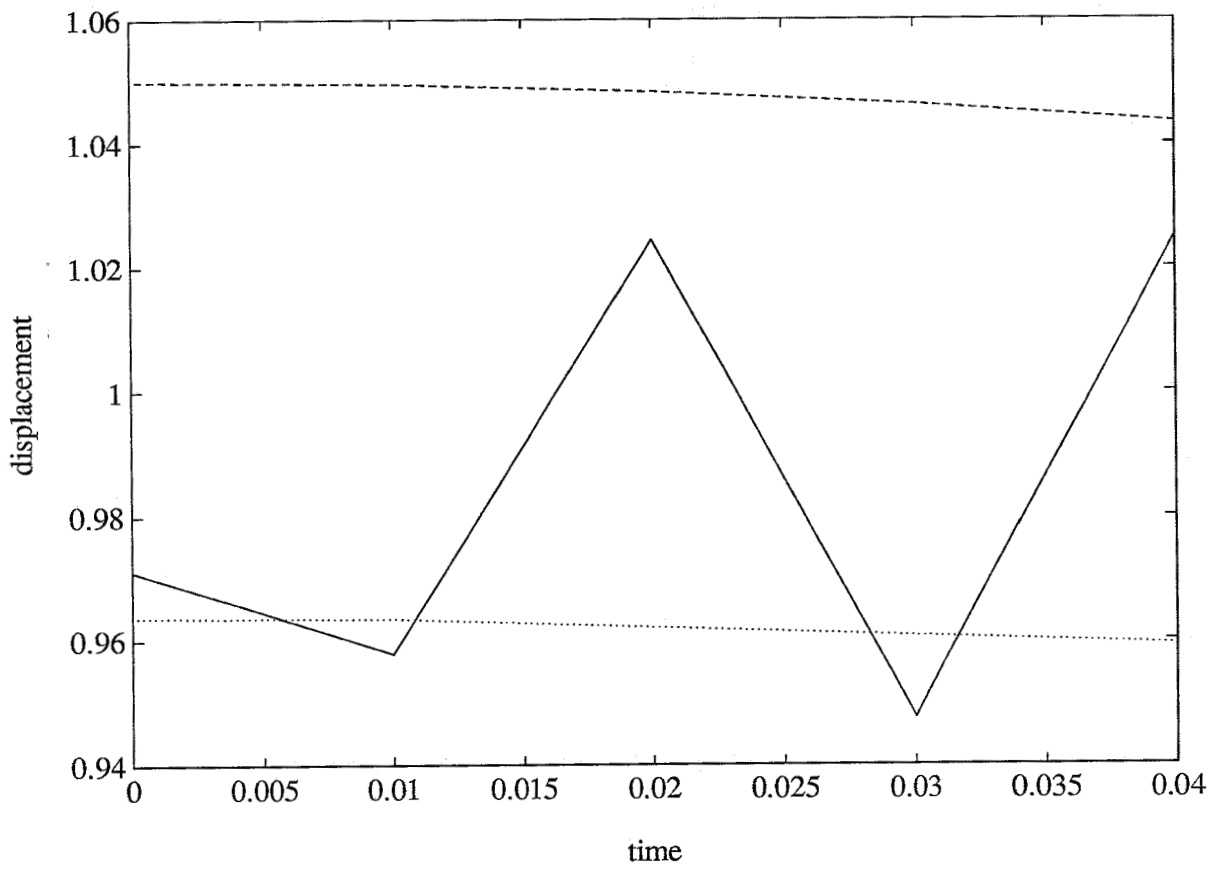

figure 5. Identification with 5 measurement points at a frequency of 100 Hz.

measurement time : 2 [s]
number of measurement points : 5
noise level : 5 %

exact parameters		estimated parameters			
1	1.0500	0.9567	0.9554	0.9552	0.9552
0	0	–0.1161	–0.0660	–0.0671	–0.0670
0.25	0.2625	0.1469	0.1322	0.1316	0.1316
7.5	7.8750	7.1842	7.2577	7.2547	7.2548

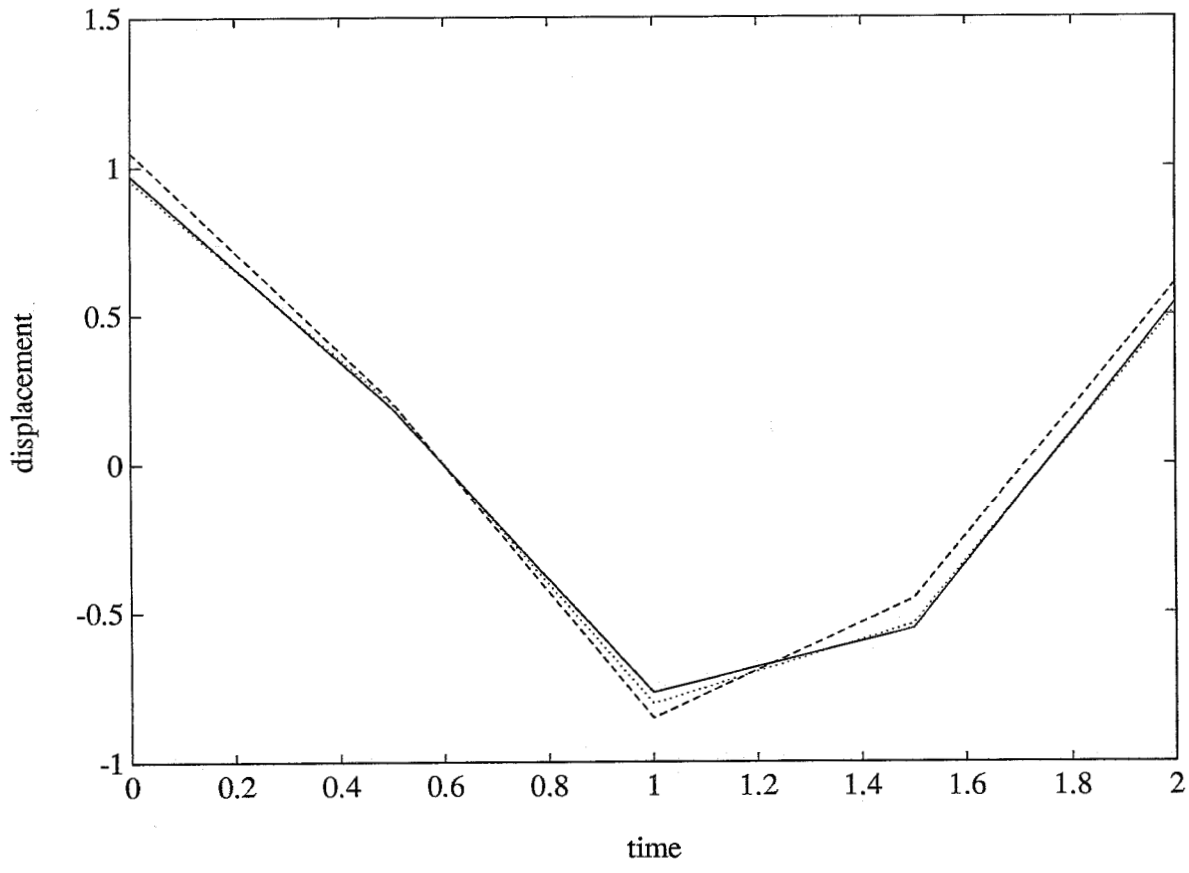

figure 6. Identification with 5 measurement points at a frequency of 2 Hz.

measurement time : 4 [s]
number of measurement points : 5
noise level : 5 %

exact parameters		estimated parameters			
1	1.0500	1.0324	1.0322	1.0322	1.0322
0	0	0.5069	0.4118	0.4164	0.4160
0.25	0.2625	0.1943	0.2033	0.2028	0.2028
7.5	7.8750	7.6997	7.7094	7.7124	7.7123

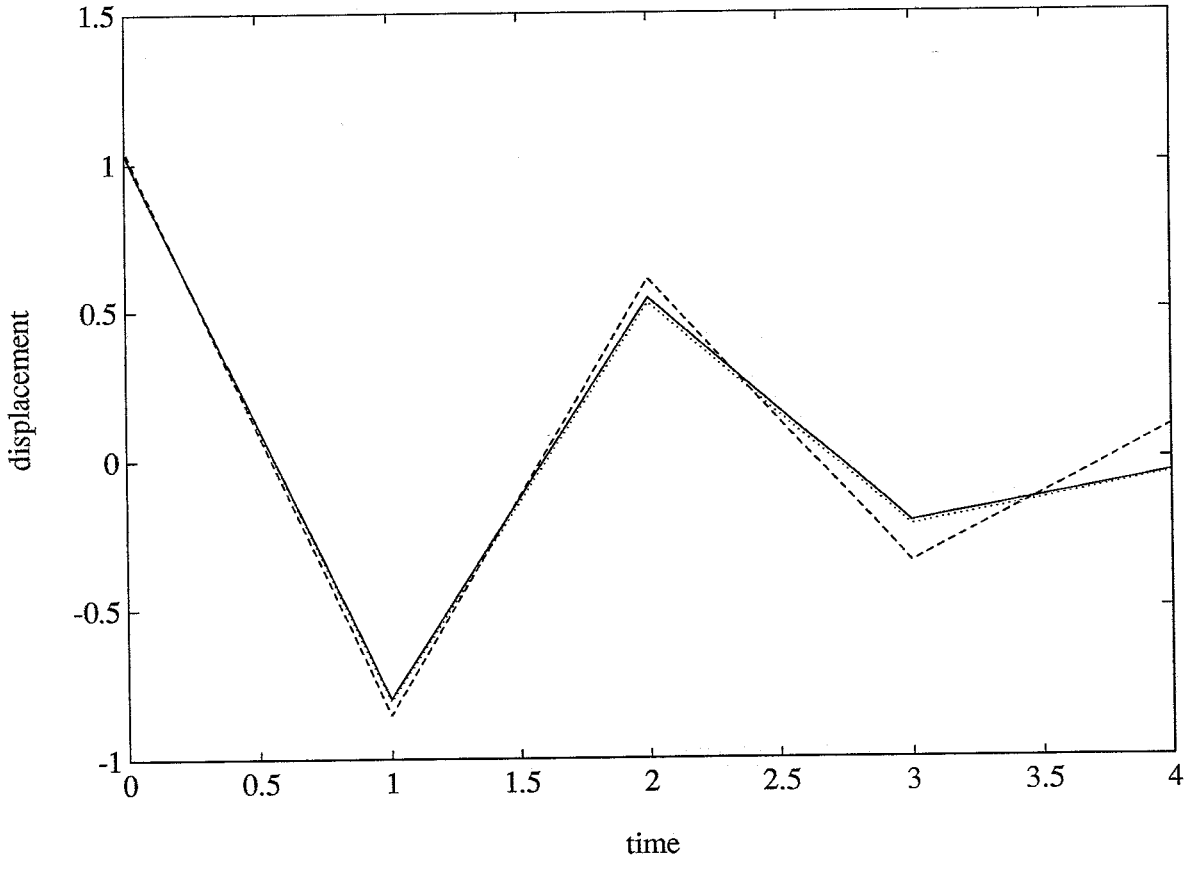

figure 7. Identification with 5 measurement points at a frequency of 1 Hz.

The second problem, sensitivity to scaling, is a topic that is known in optimization. The best thing to do is to derive another, if possible dimensionless, form of the differential equations in order to scale the parameters to the same order of magnitude.

The third problem, numerical overflow by inadmissable parameter combinations, can be avoided by employment of under – relaxation. The update of the parameter vector is damped by a user defined factor which also slows down the convergence considerably. Probably other optimization methods can be very helpful in this case. One could think of employment of the Levenberg – Marquardt method or constrained optimization.

The last problems due to discontinuities is also a topic of research in the field of numerical optimization. It will be clear that discontinuities in the forcing functions or friction functions will result in discontinuities in the derivatives of the object functions or no derivative at all. For this reason discontinuous functions should be avoided or estimated by continuous functions.

Now we are able to define good measurements and dynamical models, we can discuss the identification of oleo's. In the sequel of this report we will compare the results of the report by Batill and simulations from a MATLAB (1989) parameter estimation program developed on an AT – PC. Both the results of time integration and periodic solution methods with unweighted least squares estimators will be discussed.

The differential equation that approximates the dynamical behaviour of the oleo is the model of Milwitzky and Cook which reads

$$\ddot{u} + \theta_1 |\dot{u}| \dot{u} + \theta_2 (1/(1-\theta_3 u))^{\theta_4} - grav = 0 \quad . \tag{6.3}$$

A typical response of this system, according to the realistic parameters chosen by Batill, is shown in figure 8. All equations and derivatives used in the parameter estimation program can be found in appendix A.

Parameter	Value
u(0)	0.40 feet
$\dot{u}(0)$	0.00 feet/s
θ_1	5.08 feet^{-1}
θ_2	5.00 feet s^{-2}
θ_3	1.62 feet^{-1}
θ_4	1.00

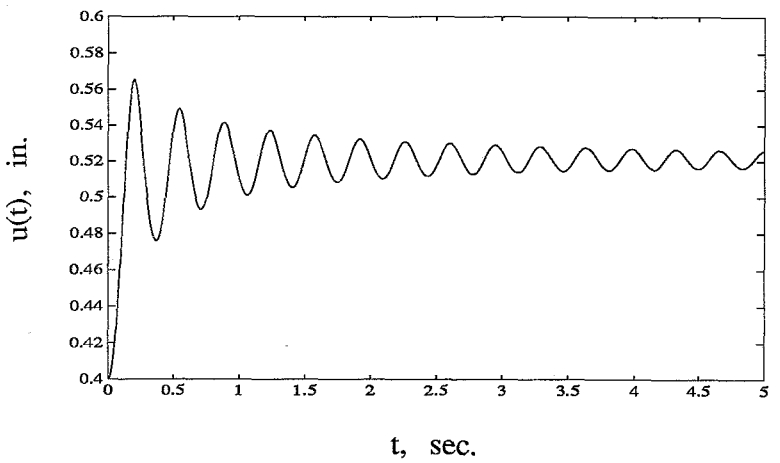

figure 8. Typical free response of an oleo

6.1 Time integration method

Comparison of Batill's results and the results of the MATLAB parameter estimation program following the procedure of section 5.1 showed no significant differences. The results of the MATLAB program are somewhat better because this program performs all computations in double precision, while Batill's code was in single precision.

For comparison of the time integration and the periodic parameter estimation program only two identifications will be discussed, both with 50 measurement points at a sample rate of 100 Hz and exact parameters equal to the parameters proposed by Batill as shown in figure 8. The first identification is performed with a displacement measurement signal that is free of noise. The identification is illustrated in figure 9. The solid line represents the exactly measured signal and, of course, the final (11^{th}) response of the exact identification process. The dashed line is the response that belongs to the intial guess of the parameter vector which was +5% in error. In this particular identification the measured signal contained sufficient information on the nonlinear system. The signal contains 1.5 cycles of motion.

measurement time : 0.5 [s]
number of measurement points : 51
noise level : 0 %

exact parameters	estimated parameters					
0.4	0.4200	0.4033	0.4002	0.3999	0.4000	0.4001
0	0	–0.1563	–0.0186	0.0037	–0.0057	–0.0041
5.08	5.3340	10.0293	5.0036	5.0959	4.9641	5.0472
5	5.2500	3.0373	3.8377	4.9767	4.8960	4.9530
1.62	1.7010	1.5533	1.5410	1.6366	1.6262	1.6202
1	1.05	1.3369	1.3240	0.9200	0.9995	1.0046
	0.4000	0.4000	0.4000	0.4000	0.4000	0.4000
	–0.0011	–0.0003	–0.0001	–0.0000	–0.0000	–0.0000
	5.0668	5.0754	5.0786	5.0796	5.0799	5.0800
	4.9971	5.0010	5.0005	5.0002	5.0001	5.0000
	1.6202	1.6201	1.6200	1.6200	1.6200	1.6200
	0.9998	0.9997	0.9999	1.0000	1.0000	1.0000

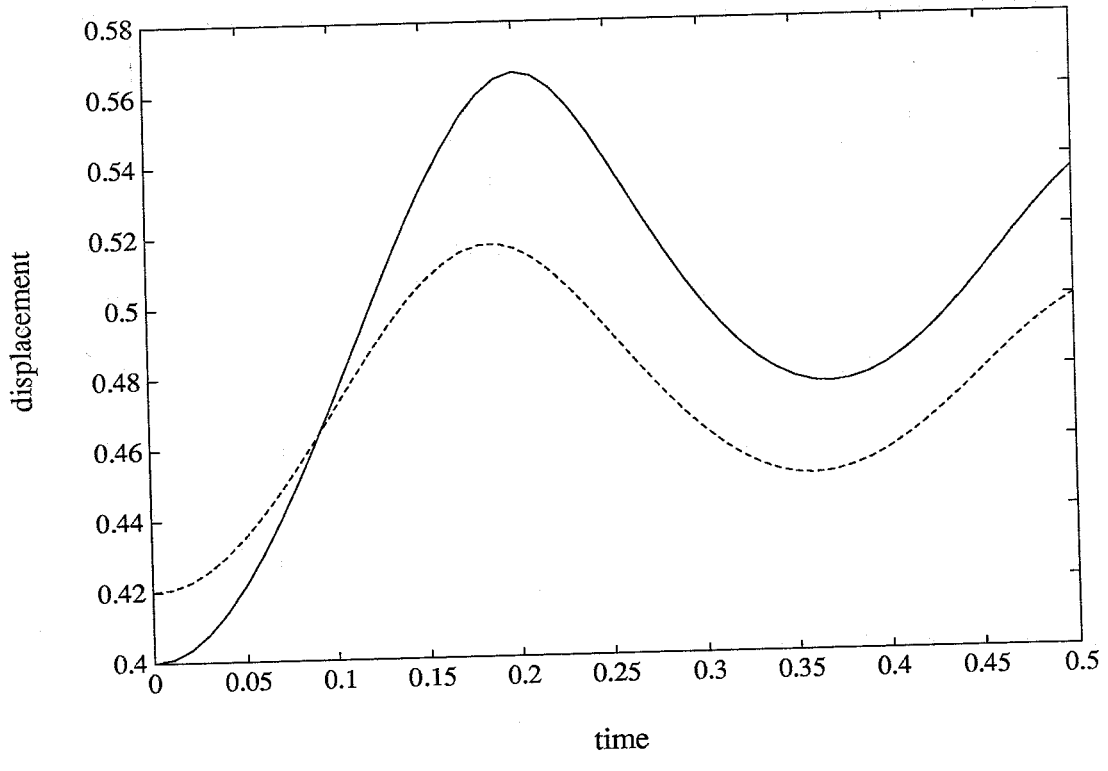

figure 9. Identification of an oleo with noise free measurements.

measurement time : 0.5 [s]
number of measurement points : 51
noise level : 1 %
under – relaxation factor : 4
number of under – rel. iterations : 6

exact parameters	estimated parameters					
0.4	0.4200	0.4001	0.3979	0.3971	0.3973	0.3974
0	0	–0.0489	0.0313	0.0535	0.0607	0.0556
5.08	5.3340	10.0115	5.5601	5.0534	5.1172	5.1677
5	5.2500	5.7969	3.5397	4.1847	4.5659	4.4191
1.62	1.7010	1.6926	1.5708	1.5847	1.5979	1.5885
1	1.05	0.7957	1.2026	1.1700	1.0938	1.1343
	0.3973	0.3973	0.3973	0.3973	0.3973	
	0.0558	0.0558	0.0560	0.0560	0.0560	
	5.1942	5.2065	5.2100	5.2111	5.2114	
	4.4027	4.3905	4.3885	4.3880	4.3879	
	1.5869	1.5860	1.5859	1.5859	1.5859	
	1.1406	1.1439	1.1445	1.1446	1.1446	

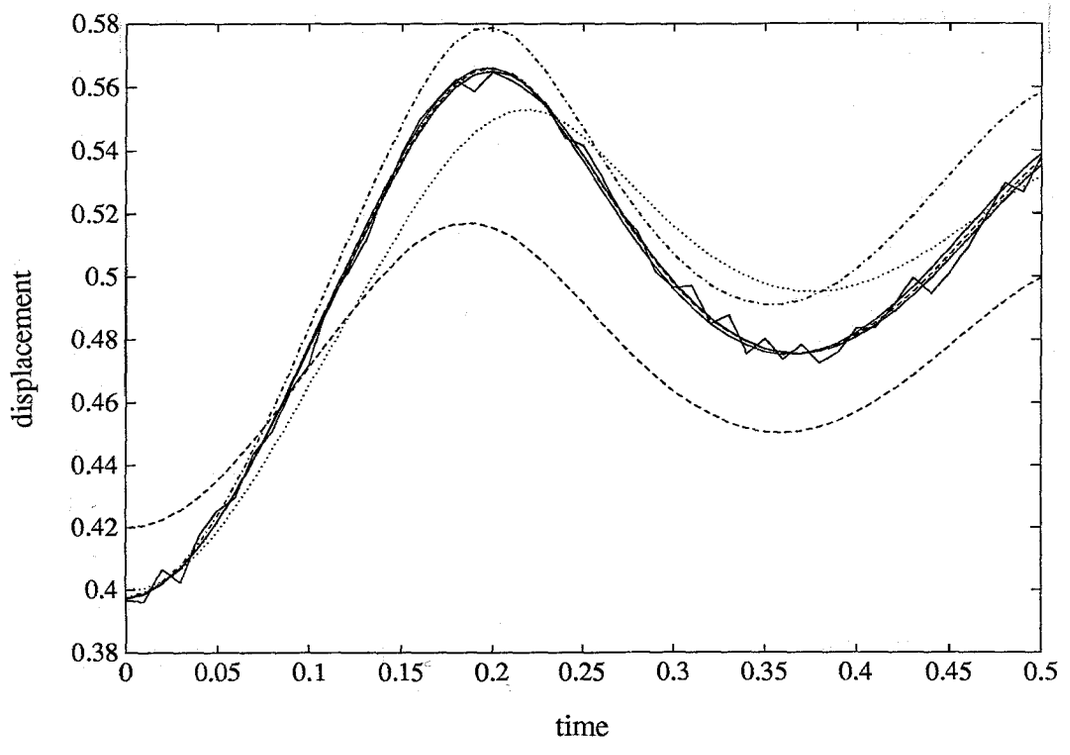

figure 10. Identification of an oleo with 1% noise level on the measurements.

The second identification is performed with a more realistic measurement signal. This time it is contaminated with 1% uniformly distributed random noise, as in Batill's article. To obtain convergence under – relaxation had to be applied with a factor of 4 over the first 6 iterations. Otherwise inadmissable parameter combinations would have caused numerical overflow. The iteration process is depicted in figure 10. Again the dashed lower line represents the initial response according to the initial guess of the parameters. The measurement signal can be recognised as a noisy solid line and the converged solution as the least square fit through the measured signal. This identification converged at last with help of under – relaxation, but the computations were troublesome and the final parameter vector is in error up to 14% . So this is not a satisfying procedure. The results could probably be improved by taking more measurements, which was limited by computer memory, and deleting the initial conditions from the estimation process as they can be measured very accurately.

6.2 Periodic solution method.

The alternative test procedure following the formulas of section 5.2 have also been implemented in a MATLAB program. As the aim of this research was to assess the new test method, a comparable identification problem had to be designed. A choise has been made for a simulated periodic response to a harmonic input signal with approximately the same amplitude as the responses in the time integration identifications. All following runs are performed with 40 measurement points. Figure 11 shows the periodic identification of the oleo. The excitation signal was a 2 Hz cosine function with an amplitude of 10 ms^{-2}. Of course, the identification converged to the exact parameters, but this time in only 3 iterations instead of 11 iterations in the time integration method.

The same simulated experiment has been performed with the 1% noise level of the comparable time integration run. Again this identification converged in only 3 iterations to parameter values that are much closer to the exact values with the damping parameter in error of 8% . The results have been added as figure 12.

As these results are promising, other simulations of the future experiments have been performed. Figure 13 shows a run in which the amplitude of the excitation force has been halved. The identification however converges to a wrong solution of the nonlinear problem. This is one of the problems that can be expected as the signals will be too small in

amplitude. In these cases the output signal will not contain sufficient information on the nonlinear system. One can see, especially in the phase plane plot, that the response resembles the ellipsoid of a linear response.

Figure 14 shows a run in which the frequency of the excitation force has been halved. The response of the system has become highly nonlinear. Nevertheless the periodic identification program converged to a reasonable solution in only 5 iterations, with maximal errors of the parameters of 4%.

The conclusion from these simulated experiments can be that the periodic identification is promising compared to identification with employment of time integration. As mentioned before a necessary condition for success is that the measurement signals contain sufficient nonlinear information on the system under consideration.

harmonic excitation frequency : 2 [Hz]
harmonic excitation force : 10
number of measurement points : 40
noise level : 0 %

exact parameters	estimated parameters			
5.08	5.3340	5.2043	5.0781	5.0800
5	5.2500	5.0037	5.0000	5.0000
1.62	1.7010	1.6186	1.6200	1.6200
1	1.0500	0.9919	1.0000	1.0000

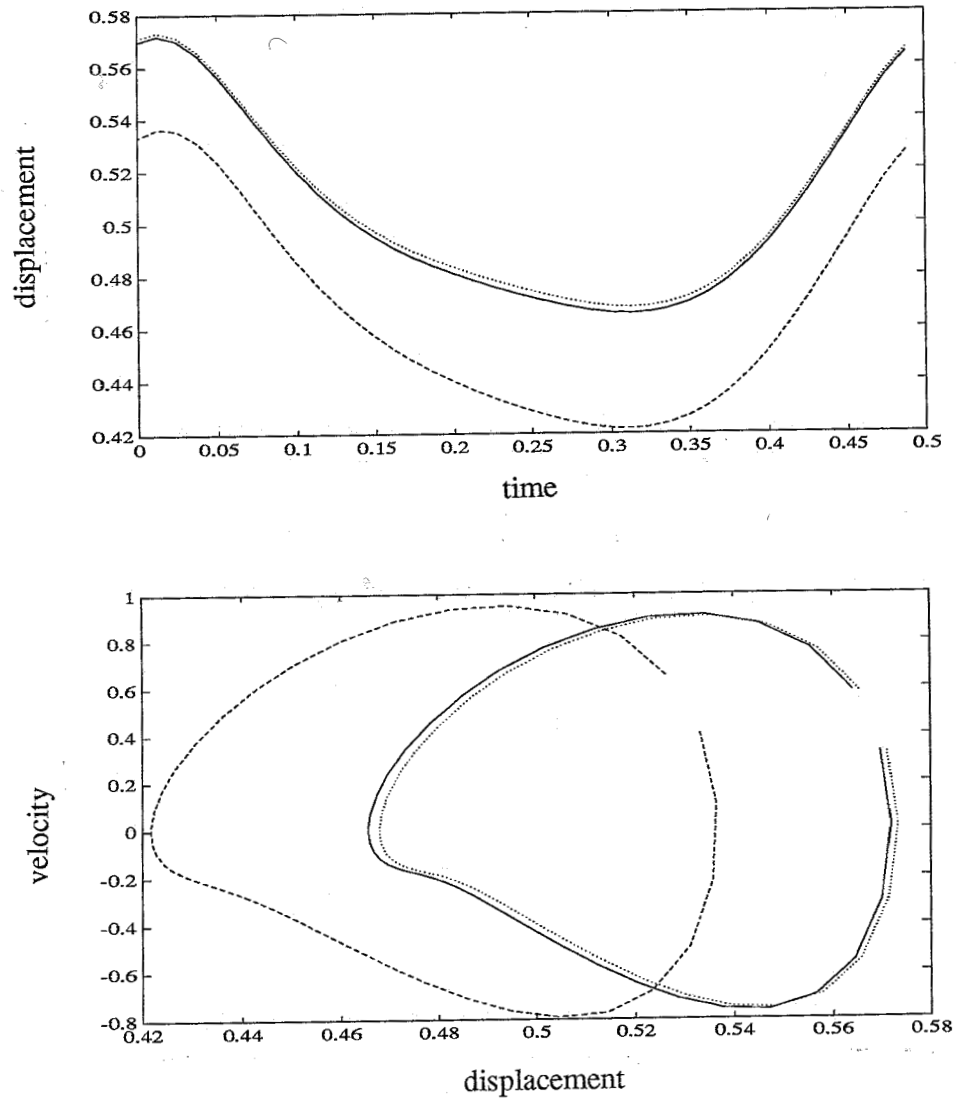

figure 11. Periodic Identification of an oleo at 2Hz with noise free measurements.

harmonic excitation frequency : 2 [Hz]
harmonic excitation force : 10
number of measurement points : 40
noise level : 1 %

exact parameters	estimated parameters			
5.08	5.3340	4.8647	4.6892	4.6911
5	5.2500	5.0390	4.9816	4.9864
1.62	1.7010	1.6196	1.6183	1.6184
1	1.0500	0.9843	1.0035	1.0029

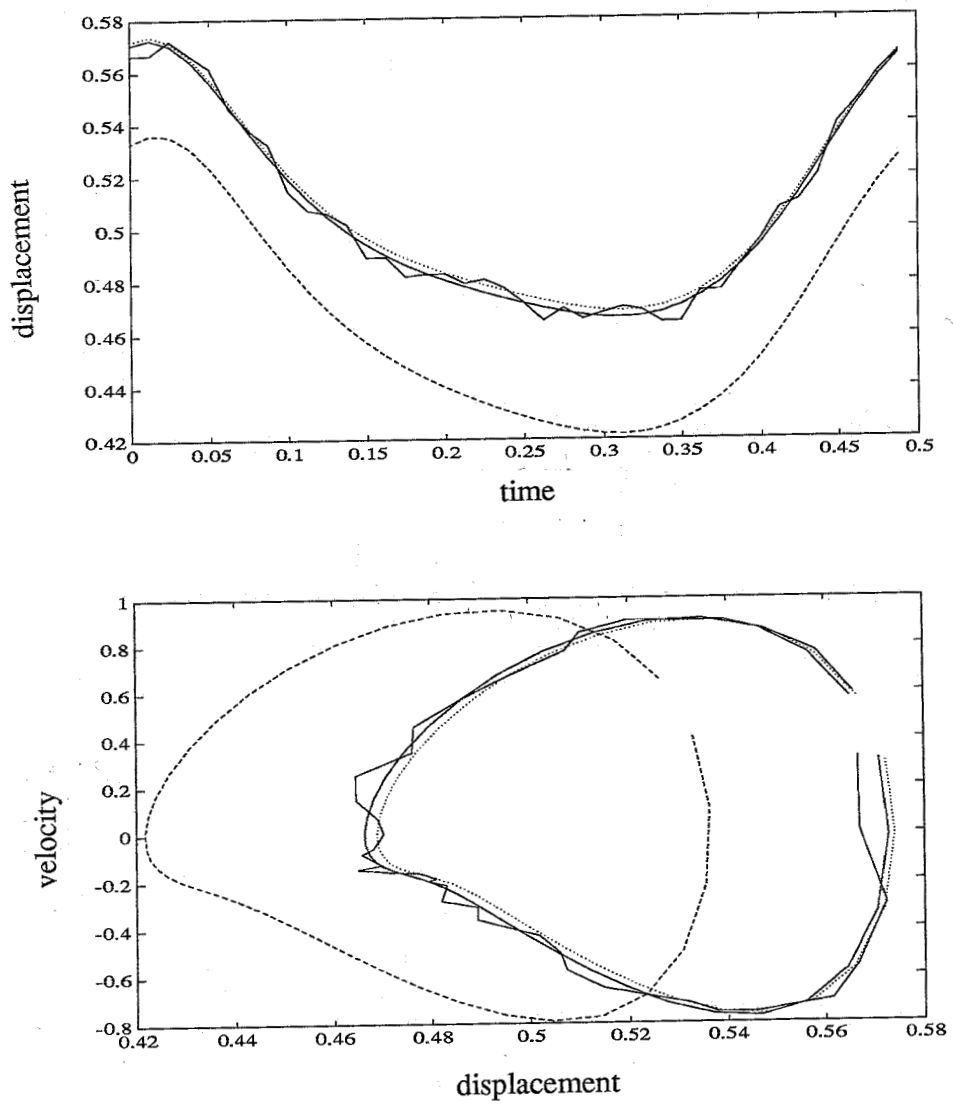

figure 12. Periodic identification of an oleo (2 Hz, 1% noise level).

harmonic excitation frequency : 2 [Hz]
harmonic excitation force : 5
number of measurement points : 40
noise level : 1 %

exact parameters	estimated parameters				
5.08	5.3340	3.7092	3.5939	3.3888	3.3689
5	5.2500	6.2261	6.4136	6.0318	6.0641
1.62	1.7010	1.6864	1.6707	1.6600	1.6607
1	1.0500	0.7403	0.7846	0.8293	0.8270

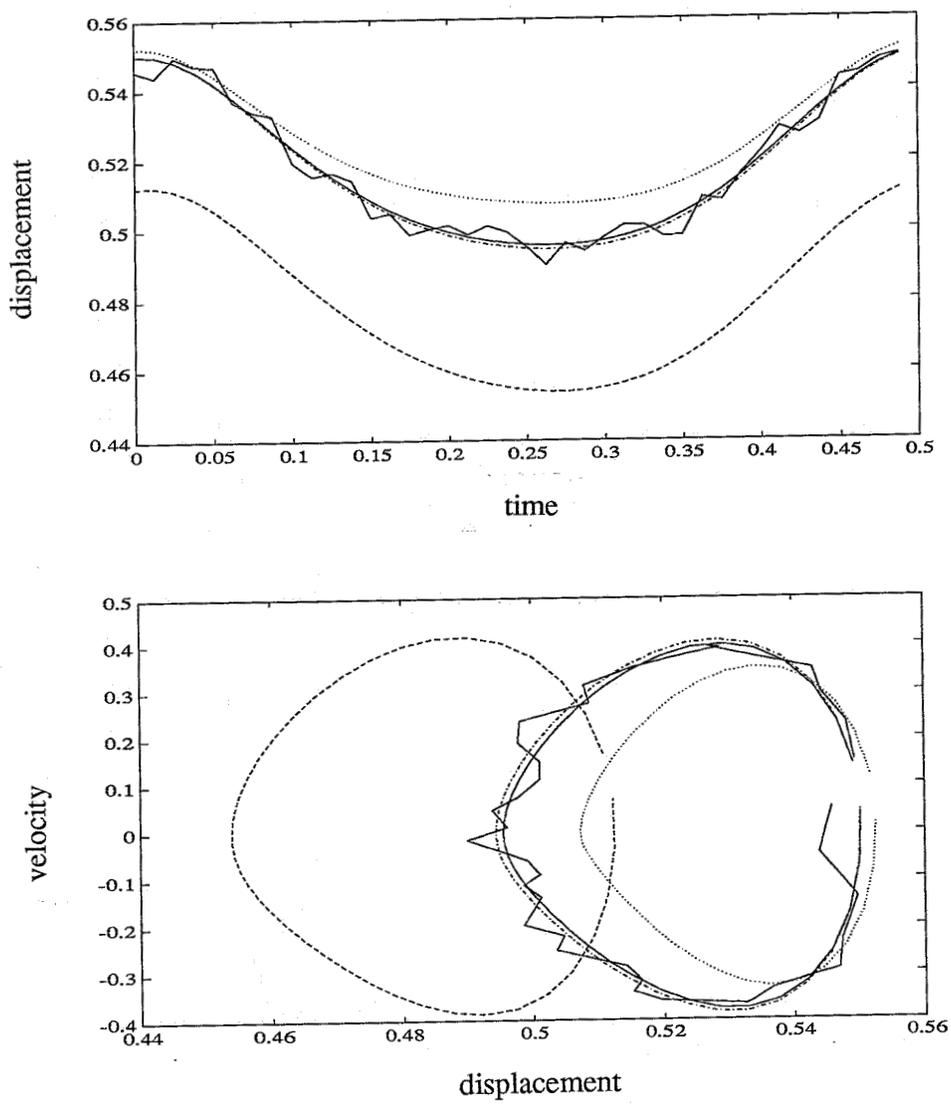

figure 13. Periodic identification of an oleo (2 Hz, 1% noise level, halved amplitude).

harmonic excitation frequency : 1 [Hz]
harmonic excitation force : 10
number of measurement points : 40
noise level : 1 %

exact parameters	estimated parameters					
5.08	5.3340	6.3019	5.2247	5.2986	5.2979	5.2979
5	5.2500	4.6829	4.7619	4.7816	4.7839	4.7838
1.62	1.7010	1.6140	1.6143	1.6164	1.6165	1.6165
1	1.0500	1.0405	1.0407	1.0344	1.0339	1.0339

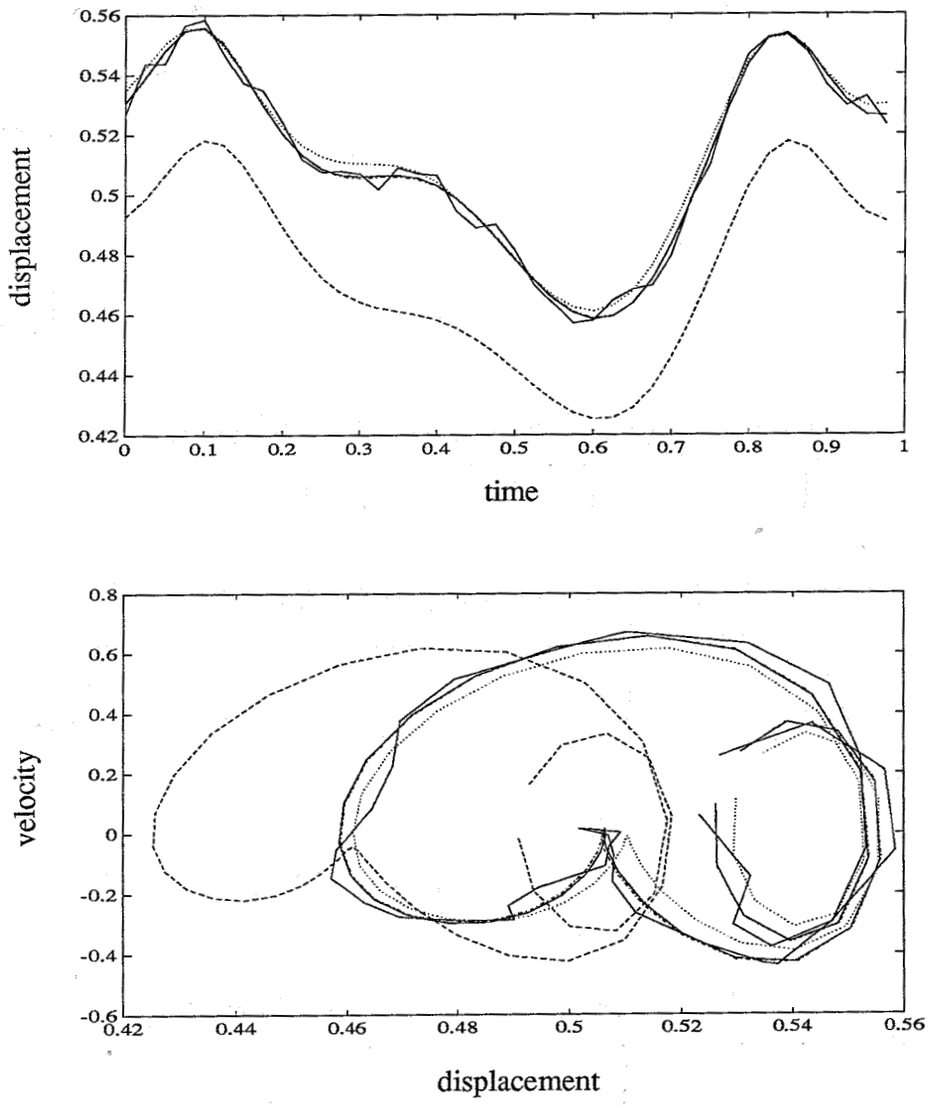

figure 14. Periodic identification of an oleo (1 Hz, 1% noise level, full amplitude).

7. CONCLUSIONS

As stated in chapter 2, a complete probabilistic model for the estimation of dynamical models consists of four elements:

- a deterministic model
- the measured data
- the probability distributions of the errors
- prior information to the unknown parameters

The deterministic model for dynamical systems can on the one hand be defined by the equations of motion or equivalent state equations or on the other hand the non – linear algebraic equations that result from the cyclic time discretisation process. It seems that the periodic approach will be numerically preferable in the estimation process because when using a time integration method we have to recompute the time response in each iteration of the estimation process from initial conditions and model equations. Using a periodic solution method enables us to compute the predicted output from the complete last known periodic time signal and model equations. So we have more information, which will speed up the convergence to a final parameter vector. The simulations of the experiments of chapter 6 confirm this statement.

The probability distribution of the errors of the measured data should be defined theoretically, by a detailed analysis but can in most cases be chosen as a normal distribution. In engineering practice most often the variance matrices of these distributions are predefined by weight matrices in least squares sums, but it is preferable and possible to estimate these variances of the measured data too by use of maximum likelihood estimators.

It also will be helpfull to add as much prior information concerning the parameters as possible. This can be done by adding relevant terms to the objective function in the case of probability density type of information or by reducing the parameter space by application of simple bounds to avoid infeasible parameter combinations.

If a normal distribution with known or unknown variances is assumed for the errors, the resulting objective function will possess favourable features for the derivatives of the objective function with respect to the parameters, because the objective function will be some sort of least squares sum. In that case the second order information can be computed very easily, by employment of Gauss – Newton methods, out of the first order derivative information. This type of second order information will only give good results if the residuals of the probabilistic model will be negligible, which is the case for good initial estimates. When not, one will have to use a modified Gauss – Newton method, as proposed in Gill, Murray & Wright (1981), which will be able to handle large residual and rank deficient problems.

For non – linear problems the computation of a variance matrix of the estimated parameters and corresponding confidence regions should be used with caution, because the computed matrix will only be of the right order of magnitude. Unfortunately, for this reason the model estimation of non – linear dynamical systems in practice will be restricted to the estimation of parameters with conservative bounds for the confidence regions.

LITERATURE

Bard, Y. (1974):
Nonlinear parameter estimation,
Academic Press, New York, ISBN 0–12–078250–2.

Batill, S.M. & Bacarro, J.M. (1988):
Modeling and identification of nonlinear dynamic systems with application to aircraft landing gear, Structures, structural dynamics and materials conference, 29th, Williamsburg, VA, april 18–20

Crooijmans, M.T.M. (1987):
On the computation of stationary deterministic behaviour of non – linear dynamic systems with applications to rotor – bearing structures, Eindhoven University of Technology, Ph.D. thesis.

Eykhoff, P. (1974):
System identification,
John Wiley & Sons, London, ISBN 0–471–24980–7.

Gill, P.E.; Murray, W. & Wright, M.H. (1981):
Practical optimization,
Academic press, London, ISBN 0–12–283952–8.

Hoogstraten, P.A.A. (1987):
Het bepalen van periodieke evenwichtsoplossingen van niet – lineaire dynamische (rotor –)systemen, Eindhoven University of Technology, report WFW – 87.036, Appendices report WFW – 87.036A, (in Dutch).

Junkins, J.L. (1978):
An introduction to optimal estimation of dynamical systems,
Sijthoff & Noordhoff, Alphen aan den Rijn, ISBN 90–286–0067–1.

Kok, J.J. (1986):
Werktuigkundige regeltechniek II,
Einhoven University of Technology, college notes, nr. 4592.

Molengraft, M.J.G. v/d (1990):
Identification of non – linear mechanical systems, for control applications,
Eindhoven University of Technology, Ph. D. thesis.

MATLAB User's Guide (1989):
The MathWorks Inc., South Natick, MA.

Verbeek, G. (1989):
Dynamics of aircraft landing and landing gear, a literature survey,
Report 89 – 454, DAF Special Products, Eindhoven, Netherlands

Young, D.W. (1986):
Aircraft landing gears – the past, present and future,
Proc. Inst. Mech. Eng. Part D, 200, no. D2, p. 75 – 92.

APPENDIX

The differential equation that approximates the dynamical behaviour of the oleo is the model of Milwitzky and Cook which reads

$$g = \ddot{u} + \theta_1 |\dot{u}| \dot{u} + \theta_2 (1/(1-\theta_3 u))^{\theta_4} - grav = 0 \quad . \tag{A.1}$$

In the periodic identification program use is made of several derivatives of this equation, which can be derived analytically

$$g_{,u} = \theta_2 \theta_3 \theta_4 (1 - \theta_3 u)^{-\theta_4} - 1 \tag{A.2}$$

$$g_{,\dot{u}} = 2\theta_1 |\dot{u}| \tag{A.3}$$

$$g_{,\ddot{u}} = 1 \tag{A.4}$$

$$g_{,\theta} = [\ \dot{u}|\dot{u}|,\ (1-\theta_3 u)^{\theta_4},\ u\theta_2\theta_4(1-\theta_3 u)^{\theta_4},$$

$$\theta_2(1-\theta_3 u)^{-\theta_4} \ln(1/(1-\theta_3 u))\ ] \tag{A.5}$$

The output equations contain only the measurement of displacement u

$$y = f(u,\theta) = u \tag{A.6}$$

The demanded derivatives on the output equations are

$$f_{,\theta} = [\ 0, 0, 0, 0\ ] \tag{A.7}$$

$$f_{,u} = 1 \tag{A.8}$$