

# Identification of parameters in large scale physical model structures, for the purpose of model-based operations

Paul M.J. Van den Hof, Jorn F.M. Van Doren and Sippe G. Douma

**Abstract** When first principles models are used for model-based operations as monitoring, control and optimization, the estimation of accurate physical parameters is important in particular when the underlying dynamical model is nonlinear. If the models are the result of partial differential equations being discretized, they are often large-scale in terms of number of states and possibly also number of parameters. Estimating a large number of parameters from measurement data leads to problems of identifiability, and consequently to inaccurate identification results. The question whether a physical model structure is identifiable, is usually considered in a qualitative way, i.e. it is answered with a yes/no answer. However since also nearly unidentifiable model structures lead to poor parameter estimates, the question is addressed how the model structure can be approximated so as to achieve local identifiability, while retaining the interpretation of the physical parameters. Appropriate attention is also given to the relevant scaling of parameters. The problem is addressed in a prediction error setting, showing the relation with gradient-type optimization algorithms as well as with Bayesian parameter estimation.

---

Paul M.J. Van den Hof  
Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD  
Delft, The Netherlands, e-mail: p.m.j.vandenhof@tudelft.nl

Jorn F.M. Van Doren  
Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD  
Delft, The Netherlands, e-mail: j.f.m.vandoren@tudelft.nl and Shell International E&P, Kessler  
Park 1, 2288 GS Rijswijk, The Netherlands, e-mail: jorn.vandoren@shell.com

Sippe G. Douma  
Shell International Exploration and Production, P.O. Box 60, 2280 AB Rijswijk, The Netherlands,  
e-mail: sippe.douma@shell.com

## 1 Introduction

Complex dynamical physical processes raise many challenges for model-based monitoring, control and optimization. On-line reconstruction of non-measurable variables, design of appropriate feedforward and feedback control strategies, as well as economic optimization of processes under appropriate operational constraints, generally require the availability of a reliable process model, preferably accompanied by a quantification of its reliability (uncertainty). If the dynamics of the considered process is linear, then a process model can be obtained by applying black-box system identification, which provides a well-studied set of tools for identifying linear models on the basis of experimental data [15, 21]. If there is a particular interest in the identification of physical parameters, this often does not raise any additional problems: one has to choose the right (physics-based) model structure and identify the parameters through one of the available (possibly non-convex) optimization methods. The only issue that has to be taken care of is that the physical model structure is identifiable, implying that the several physical parameters can be distinguished from each other on the basis of the model's input-output behavior.

In the situation that the process dynamics is nonlinear, it can often be linearized around an operating point (as e.g. in continuous-type industrial/chemical production processes) and the above mentioned linear approach can be followed leading to a linear (approximate) model. However when essential nonlinear dynamical phenomena are involved and the user needs to capture this dynamics in the model, it is much harder to come up with generic black-box techniques for identification. Although there are interesting attempts to capture the nonlinear phenomena in (black-box) model structures as Wiener and/or Hammerstein models, [1, 33], and linear parameter-varying (LPV) models [14, 29, 2, 25, 24, 32, 34], information on the underlying physical structure of the nonlinearities is very often required for selection of an appropriate model structure.

In some processes it is desirable to capture the real underlying nonlinear dynamic structure of the process in order to make reliable long-term predictions. First-principles model then provide the structure of the model, while incorporated (physical) parameters have to be estimated on the basis of experimental data. Especially in situations where the first principles models are given by partial differential equations (pde's), the required step of discretizing the equations in space and time generally leads to complex models with a large number of states and possibly also a large number of unknown (physical) parameters. For an interesting example of this situation in a problem of (oil) reservoir engineering, the reader is referred to [13], where an industrial example of optimal oil recovery from reservoirs over the life time of the reservoir (possibly  $> 20$  years) is shown in the form of a nonlinear (batch-type) process with a number of states and parameters exceeding the order of  $10^5$ .

Identifying extremely large number of parameters from measurement data leads to serious problems, and at least it leads to the question which model properties can be reliably estimated from the available measurement data. From a model-based operations point of view (monitoring, control, optimization) it makes sense to limit the complexity of an identified model to a level where the model can be reliably vali-

dated from data. If not, the parameter estimates might be highly determined by the -random- experiment that is done (overfit) leading to unreliable model predictions. In identification this problem is addressed by the notion of identifiability (of a model structure), and directly coupled to the variance of estimated parameters.

In this chapter the notion of identifiability will be evaluated in the scope of the high-complexity type of processes discussed above. Our argument will be that having an (locally) identifiable model structure will not be sufficient to provide reliable parameter estimates in large scale physical model structures. Methods will be presented that allow to reduce the parameter space to limited dimension, while being able to reliably estimate the reduced parameters and maintaining their physical interpretation. To this end the qualitative notion of (local) identifiability (with a yes/no answer) is generalized to a quantitative notion, removing that parameter subspace from the parametrization that can only be estimated with excessively large variance.

## 2 Identifiability - the starting point

The notion of identifiability refers -roughly speaking- to the question whether parameter changes in the model can be observed in the model output signal (output identifiability) or in the model's input-output map or transfer function (structural identifiability).

The notion of output identifiability has been studied in e.g. [11] and [15]. The notion of structural identifiability was first stated by [3] and has been extensively studied in the field of compartmental modeling ([10], [20]). State-space model parameterizations have been analyzed by [9] and [30]. Lately there has been a renewed interest in structural identifiability analysis, with contributions from [22] and [27]. In its essence, identifiability properties are global properties, i.e. holding for the full parameter space, as e.g. considered in [16] and [7] where parameter mappings are studied to analyse global identifiability. However restricting attention to a local analysis is often the only situation that is feasible in terms of computational complexity. As a result we will focus on local properties in this chapter only.

Consider a nonlinear dynamical model that generates output predictions according to<sup>1</sup>:

$$\hat{\mathbf{y}} = h(\mathbf{u}, \boldsymbol{\theta}; x_0), \quad (1)$$

where  $\hat{\mathbf{y}}$  is a prediction of  $\mathbf{y} := [y_1^T \ \dots \ y_N^T]^T$  denoting output signal measurements  $y_k \in \mathbb{R}^p$  stacked over time,  $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$  the parameter vector,  $\mathbf{u} := [u_1^T \ \dots \ u_N^T]^T$  the input vector  $u_k \in \mathbb{R}^m$  stacked over time, and  $x_0$  the initial state vector. Since the model (1) is parameterized it represents an input/output model structure.

The definition of local identifiability now is given as follows([11]):

**Definition 1** *An input/output model structure  $h(\boldsymbol{\theta}, \mathbf{u}; x_0) : \Theta \rightarrow \mathcal{H}$  is called locally identifiable in  $\boldsymbol{\theta}_m \in \Theta$  for a given  $\mathbf{u}$  and  $x_0$ , if for all  $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$  in the neighborhood of*

<sup>1</sup> Without loss of generality we restrict attention to predictors that are not dependent on output measurements  $y$ , which in an LTI-setting is referred to as Output Error predictors.

$\theta_m$  holds that

$$\{h(\mathbf{u}, \theta_1; x_0) = h(\mathbf{u}, \theta_2; x_0)\} \Rightarrow \theta_1 = \theta_2.$$

If we linearize the nonlinear process dynamics around a chosen operating point or trajectory, a linear dynamical system results. This system can be modelled by an LTI input-output model, represented by the transfer function  $G$ , leading to an output predictor

$$\hat{y}_k = G(q, \theta)u_k,$$

with  $q$  the shift operator  $qu_k = u_{k+1}$ .

The notion of local structural identifiability ([9]) is then defined by considering the properties of the parameterized transfer function  $G(q, \theta)$ :

**Definition 2** An input/output model structure  $G : \Theta \rightarrow \mathcal{G}$  with  $\Theta \subset \mathbb{R}^q$  and  $\mathcal{G} \subset \mathbb{R}(z)^{p \times m}$  is called locally structural identifiable in  $\theta_m \in \Theta$  if for all  $\theta_1, \theta_2$  in the neighborhood of  $\theta_m$  holds that

$$\{G(z, \theta_1) = G(z, \theta_2)\} \Rightarrow \theta_1 = \theta_2. \quad (2)$$

Here  $\mathbb{R}(z)^{p \times m}$  is the set of all  $p \times m$  rational transfer function matrices in the complex indeterminate  $z$ .

Note that in contrast with (1) this latter notion is not dependent on either an input signal nor an initial state. Structural identifiability will be considered in section 8, where a link is made between structural identifiability and identifiability.

In general identifiability questions are considered qualitatively, i.e. deciding whether a model structure is either identifiable or not. The tests required for this evaluation are typically rank evaluations of matrices, as e.g. Fisher's information matrix, around a particular local operating point in the parameter space, see e.g. [6]. However, when considering parameters in large scale (nonlinear) physical models it is relevant to raise the question how the notion of identifiability can be quantified. This implies addressing the question which part of the parameter space is best identifiable, and which part of the model structure can be approximated so as to achieve local identifiability, while retaining the interpretation of the physical parameters. For structural identifiability this question was preliminary addressed in [27]. In [4] the degree of identifiability was introduced. In [26] principal component analysis was applied to determine which parameters can be identified. Assessing identifiability can also be done a posteriori, after the identification of all parameters, by evaluating the parameter variance, see e.g. [12].

In this chapter we will further investigate how the notions of identifiability can be quantified to allow for a reduction in the parameter space with physically interpretable parameters.

### 3 Testing local identifiability in identification

#### 3.1 Introduction

In a model identification framework we consider parameter estimation methods that are characterized by minimizing a cost function  $V(\theta)$ :

$$V(\theta) := \frac{1}{2} \boldsymbol{\varepsilon}(\theta)^T P_v^{-1} \boldsymbol{\varepsilon}(\theta), \quad (3)$$

where the prediction error sequence  $\boldsymbol{\varepsilon}$  is defined as

$$\boldsymbol{\varepsilon}(\theta) = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - h(\theta, \mathbf{u}; x_0), \quad (4)$$

where  $\mathbf{y}$  denotes the measured output sequence and  $\hat{\mathbf{y}}$  the predictor sequence, and  $P_v$  is a weighting matrix that could represent (an estimate of) the covariance matrix of the noise sequence  $\mathbf{v}$  that is supposed to act on the measured output. In the rest of the chapter the shorthand notation  $\hat{\mathbf{y}}(\theta)$  is used to indicate  $h(\mathbf{u}, \theta; x_0)$ .

The Jacobian of  $V(\theta)$  with respect to the parameters is

$$\frac{\partial V(\theta)}{\partial \theta} = \frac{\partial \boldsymbol{\varepsilon}(\theta)^T}{\partial \theta} P_v^{-1} \boldsymbol{\varepsilon}(\theta) = - \frac{\partial \hat{\mathbf{y}}(\theta)^T}{\partial \theta} P_v^{-1} (\mathbf{y} - \hat{\mathbf{y}}(\theta)). \quad (5)$$

The Hessian of  $V(\theta)$  with respect to the parameters is

$$\frac{\partial^2 V(\theta)}{\partial \theta^2} = \frac{\partial \boldsymbol{\varepsilon}(\theta)^T}{\partial \theta} P_v^{-1} \left( \frac{\partial \boldsymbol{\varepsilon}(\theta)^T}{\partial \theta} \right)^T + S = \frac{\partial \hat{\mathbf{y}}(\theta)^T}{\partial \theta} P_v^{-1} \left( \frac{\partial \hat{\mathbf{y}}(\theta)^T}{\partial \theta} \right)^T + S, \quad (6)$$

where  $S$  denotes the second-order information in  $\frac{\partial^2 V(\theta)}{\partial \theta^2}$ . The Jacobian and Hessian are for a given  $\theta$  and a given operating point (given by  $\mathbf{u}$  and  $x_0$ ). Parameter estimation now consists in finding the parameter estimate as a minimizing argument of the cost function  $V(\theta)$

$$\hat{\theta} := \arg \min_{\theta} V(\theta). \quad (7)$$

At  $\hat{\theta}$  the cost function  $V(\theta)$  is minimized and the Jacobian (5) at  $\hat{\theta}$  is zero, i.e.  $\frac{\partial V(\theta)}{\partial \theta} = 0$  at  $\hat{\theta}$ .

#### 3.2 Analyzing local identifiability in $\hat{\theta}$

Local identifiability in  $\hat{\theta}$  is generally evaluated by the test whether the optimization problem (7) has a unique solution in the parameter space. By locally approximating the cost function  $V(\theta)$  by a quadratic function<sup>2</sup> (and thus neglecting the second

<sup>2</sup> This is achieved by approximating  $\hat{\mathbf{y}}(\theta)$  with a first-order Taylor expansion around  $\hat{\theta}$ .

order term  $S$  in (6)), uniqueness of  $\hat{\theta}$  is guaranteed if the Hessian at  $\hat{\theta}$  is positive definite, i.e.  $\frac{\partial^2 V(\theta)}{\partial \theta^2} > 0$  at  $\hat{\theta}$ , which in this case is equivalent to  $\text{rank} \frac{\partial^2 V}{\partial \theta^2} = q$ . This is a sufficient condition for local identifiability in  $\hat{\theta}$ , see e.g. [3], [9] and [15].

The considered rank test is naturally performed by applying a singular value decomposition (SVD):

$$\frac{\partial^2 V(\theta)}{\partial \theta^2} = U \Sigma V^T = [U_1 \ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix},$$

where matrices  $U$  and  $V$  are unitary matrices,  $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_p)$  with  $\sigma_1 \geq \dots \geq \sigma_p$ .

If  $p = q$  then identifiability is confirmed. If  $p < q$  then the column space of  $U_1$  represents the subspace of the parameter space that is identifiable, and the column space of  $U_2$  is its orthogonal complement, characterizing the subspace that is not identifiable.

As a result, the SVD of the Hessian can be used to extend the qualitative treatment of the question whether or not a particular model structure is identifiable, to a quantitative property of specifying the identifiable parameter space. The columns of  $U_1$  basically act as basis functions in the parameter space, determining the linear combinations of the original parameters that will be identifiable from the measurements. Differently formulated, this would point to a reparametrization of the model structure by defining a reduced order parameter  $\rho \in \mathbb{R}^p$  defined by

$$\theta = U_1 \rho \tag{8}$$

leading to an identifiable model structure in the parameter  $\rho$ . The attractive feature of this mapping is that it allows to identify  $\rho$  while the estimated result  $\hat{\rho}$  can be uniquely interpreted in terms of the original physical parameters  $\theta$  through the mapping (8). The limitation of the approach is of course that only *linear* parameter transformations are considered.

### 3.3 Approximating the identifiable parameter space

When in the SVD of the Hessian singular values are found that are (very) small, this points to directions in the parameter space that have very limited (but nonzero) influence on the cost function  $V$ . In identification terms this correspond to directions in the parameter space in which the variance is (very) large. The Hessian evaluated at  $\hat{\theta}$  is connected to the variance of  $\hat{\theta}$ , since for the Gaussian case (and provided that  $\hat{\theta}$  is a consistent estimate) it follows that

$$\text{cov}(\hat{\theta}) = J^{-1}$$

with  $J$  the Fisher information matrix

$$J = \mathbb{E} \left[ \left. \frac{\partial^2 V(\theta)}{\partial \theta^2} \right|_{\hat{\theta}} \right], \quad (9)$$

where  $\mathbb{E}$  denotes expectation ([15]).

We are interested in specifying that part of the parameter space that is best identifiable by removing the subspace that has only a very small influence on the cost function  $V$ . This reasoning would point to removing those parameter (combinations) from the model structure for which the variance is very large, as was also addressed in [26] for nonlinear parameter mappings, and in [17] for single parameters.

The essential information on the SVD of the Hessian is now obtained from:

$$\frac{\partial \hat{y}(\theta)^T}{\partial \theta} P_v^{-\frac{1}{2}} = [U_1 \ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \quad (10)$$

where the separation between  $\Sigma_1$  and  $\Sigma_2$  is chosen in such a way that the singular values in  $\Sigma_2$  are considerably smaller than those in  $\Sigma_1$ .

If we now reparametrize the model structure by employing the reduced parameter  $\rho$  determined by  $\theta = U_1 \rho$ , we have realized a model structure approximation, in which the parameters to be identified are well identifiable with a limited variance and the physical interpretation of the parameters remains untouched. The singular vectors that occur as the columns in  $U_1$  actually can be seen as basis functions in the parameter space.

With the SVD (10) it follows that the sample estimate of the covariance matrix of  $\hat{\theta}$  becomes:

$$\text{cov}(\hat{\theta}) = \begin{cases} [U_1 \ U_2] \begin{bmatrix} \Sigma_1^{-2} & 0 \\ 0 & \Sigma_2^{-2} \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} & \text{for } \text{trace}(\Sigma_2) > 0 \\ U_1 \Sigma_1^{-2} U_1^T & \text{for } \Sigma_2 = 0 \end{cases} \quad (11)$$

while the sample estimate of the covariance matrix of the reparametrized parameter estimate  $U_1 \hat{\rho}$  is given by

$$\text{cov}(U_1 \hat{\rho}) = U_1 \Sigma_1^{-2} U_1^T. \quad (12)$$

This shows that if  $\Sigma_2 = 0$  there is no benefit of the reparametrization in terms of variance of the estimated parameter  $\hat{\theta}$ . However if nonzero singular values are discarded in  $\Sigma_2$ , i.e. if  $\text{trace}(\Sigma_2) > 0$  then

$$\text{cov}(\hat{\theta}) > \text{cov}(U_1 \hat{\rho})$$

showing a covariance that is reduced by the reparametrization. This reduction is particularly interesting if  $\Sigma_2$  contains a (very) large number of small singular values.

## 4 Parameter scaling in identifiability

The notions of identifiability are defined in such a way that the result is independent of any particular scaling of parameters. A scaling happens when choosing a particular physical unit for a particular parameter, as e.g. using either [nm] or [m] as measure of distance. One would also expect that issues around identifiability should not be dependent on these scalings. Also the analysis and test in section (3.2) is independent of parameter scaling; a scaling of parameters leads to scaled singular values, but nonzero singular values remain nonzero after scaling and vice versa. However when considering the singular value decomposition in the approximating situation of section (3.3) parameter scaling does have an influence on the numerical values that occur in  $\Sigma_1, \Sigma_2$ , and therefore can essentially influence the choice for separating  $\Sigma_1$  and  $\Sigma_2$ . This particularly plays a role when the physical parameters contain different physical quantities. E.g. in the case of [13] the physical parameters relate to saturations (oil/water percentages) and permeabilities in each separate grid-block that is a result of spatial discretization. The underlying question of parameter scaling is then: how to balance the variability in the different physical parameters.

It appears that in the approach presented above the absolute variance of parameters is used as a measure for selection, and as a result the selected parameter space will become dependent on the chosen parameters scales/units. If it is preferred to arrive at a selection mechanism that is scaling independent, the relative variance of parameters is an attractive choice, i.e.

$$\text{cov}(\Gamma_{\hat{\theta}}^{-1} \hat{\theta})$$

where  $\Gamma_{\hat{\theta}} = \text{diag}(|\hat{\theta}_1| \dots |\hat{\theta}_q|)$ . This motivates the analysis of a scaled Hessian

$$\Gamma_{\hat{\theta}} \frac{\partial^2 V(\theta)}{\partial \theta^2} \Big|_{\hat{\theta}} \Gamma_{\hat{\theta}}, \quad (13)$$

related to the scaled Fisher information matrix  $\tilde{J}$ :

$$\tilde{J} = \mathbb{E} \left[ \Gamma_{\hat{\theta}} \frac{\partial^2 V(\theta)}{\partial \theta^2} \Big|_{\hat{\theta}} \Gamma_{\hat{\theta}} \right]. \quad (14)$$

The appropriate selection of the identifiable parameter space is then obtained by applying an SVD according to:

$$\Gamma_{\hat{\theta}} \frac{\partial \hat{y}(\theta)^T}{\partial \theta} P_v^{-\frac{1}{2}} = [U_1 \ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}. \quad (15)$$

Consequences of this parameter scaling are illustrated for some simple examples in Section 9.

Note that the evaluation of the relative variance of parameter estimates for model structure selection is also done in classical methods when considering the standard deviation of an estimated parameter related to the parameter value itself, see e.g.

[15] and [12]. However usually the analysis is performed for parameters separately (e.g. is zero included in the parameter confidence interval?). In the analysis presented here linear combinations of parameters are evaluated, thus focussing on the ratio between the lengths of the principle axes of the uncertainty ellipsoids representing the parameter confidence bounds for  $\hat{\theta}$ . The parameter uncertainty ellipsoid is expressed by

$$\mathcal{D}(\alpha, \hat{\theta}) = \{\theta \mid N(\theta - \hat{\theta})^T P^{-1}(\theta - \hat{\theta}) \leq \chi_{q, \alpha}^2\}$$

with  $P = J^{-1}$  the covariance matrix of the estimator, and  $\chi_{q, \alpha}^2$  corresponds to a probability level  $\alpha$  in the  $\chi_q^2$ -distribution with  $q$  degrees of freedom. This parameter uncertainty ellipsoid is used to specify that  $\theta_0 \in \mathcal{D}(\alpha, \hat{\theta})$  with probability  $\alpha$ .

## 5 Relation with controllability and observability

In this section we will show how the identifiable parameter space that results from (10) is related to properties of controllability and observability.

We consider a strictly proper deterministic linear time-varying (LTV) model in discrete-time state-space form, that could result from linearizing a nonlinear model in the vicinity of a nominal trajectory. The model is given by

$$\begin{aligned} x_{k+1} &= A_k(\theta)x_k + B_k(\theta)u_k \\ \hat{y}_k(\theta) &= C_k(\theta)x_k, \end{aligned}$$

where subscript  $k$  denotes the time index. The sensitivity of the predicted outputs with respect to the parameter vector  $\theta$  is element-wise given by

$$\frac{\partial \hat{y}_k(\theta)}{\partial \theta(i)} = C_k(\theta) \frac{\partial x_k}{\partial \theta(i)} + \frac{\partial C_k(\theta)}{\partial \theta(i)} x_k,$$

where  $\frac{\partial x_k}{\partial \theta(i)}$  is determined by

$$\frac{\partial x_{k+1}}{\partial \theta(i)} = A_k(\theta) \frac{\partial x_k}{\partial \theta(i)} + \underbrace{\frac{\partial A_k(\theta)}{\partial \theta(i)} x_k + \frac{\partial B_k(\theta)}{\partial \theta(i)} u_k}_{:= \tilde{u}_k^{\theta(i)}}. \quad (16)$$

Without loss of generality we assume that  $\frac{\partial C_k}{\partial \theta(i)} = 0$ , since  $C_{k+1}$  can be made independent of  $\theta$  by redefining the state. Note that the effect of a parameter change is weighted by the value of current state and input, i.e. in (16)  $\frac{\partial A_k(\theta)}{\partial \theta(i)}$  is weighted by  $x_k$  and  $\frac{\partial B_k}{\partial \theta(i)}$  is weighted by  $u_k$ . This means that given a specific model structure, outputs are more sensitive to parameters associated with states that have a large value. In stacked form we can write

$$\begin{aligned}
& \left( \frac{\partial \hat{\mathbf{y}}(\theta)^T}{\partial \theta} \right)^T = \\
& = \underbrace{\begin{bmatrix} C_1 & & 0 \\ & C_2 & \\ & & \ddots \\ 0 & & & C_N \end{bmatrix}}_{\tilde{\mathcal{O}} \in \mathbb{R}^{N \times n}} \underbrace{\begin{bmatrix} I & 0 & \dots \\ A_1 & I & \\ A_2 A_1 & A_2 & I \\ \vdots & & \ddots & \vdots \\ A_{N-2} \dots A_1 & A_{N-2} \dots A_2 & \dots & A_{N-2} & I \\ A_{N-1} \dots A_1 & A_{N-1} \dots A_2 & \dots & \dots & A_{N-1} & I \end{bmatrix}}_{\tilde{\mathcal{U}} \in \mathbb{R}^{Nn \times q}} \times \begin{bmatrix} \tilde{u}_0^{\theta(1)} & \dots & \tilde{u}_0^{\theta(i)} & \dots & \tilde{u}_0^{\theta(q)} \\ \tilde{u}_1^{\theta(1)} & \dots & \tilde{u}_1^{\theta(i)} & \dots & \tilde{u}_1^{\theta(q)} \\ \vdots & & \vdots & & \vdots \\ \tilde{u}_{N-1}^{\theta(1)} & \dots & \tilde{u}_{N-1}^{\theta(i)} & \dots & \tilde{u}_{N-1}^{\theta(q)} \end{bmatrix}, \quad (17)
\end{aligned}$$

where we have defined  $\tilde{\mathcal{O}}$  and  $\tilde{\mathcal{U}}$ . For a change in the model parameters the term  $\tilde{u}_k^{\theta(i)}$  in (16) was given by  $\frac{\partial A(\theta)}{\partial \theta^{(i)}} x_k + \frac{\partial B(\theta)}{\partial \theta^{(i)}} u_k$ . With this expression, equation (17) provides an appealing interpretation of the Jacobian of the predicted outputs with respect to the parameter vector: a change in any of the parameters is translated into a perturbation of the state of the system, which propagates over time through the dynamical system to reveal the effect on the predicted outputs. As a result the Jacobian is seen to be determined by three factors: the current state and input ( $x_k, u_k$ ), secondly, the mapping from a model parameter perturbation to a state change (sensitivities of  $A$  and  $B$ ), and thirdly the mapping from a state perturbation to a change in the output (observability properties in  $\tilde{\mathcal{O}}$ ). Indeed only parameter changes that result in state perturbations contained in the column space of ( $\tilde{\mathcal{O}}$ ) can be identified. Moreover, the current state and input need to be nonzero. In the situation that the initial state  $x_0 = 0$ , and  $\frac{\partial B}{\partial \theta} u = 0$ , it follows that the state will be significant only in the controllable state space. Note, however, that the initial state  $x_0$  and natural disturbances of the states can also contribute to  $x$  being nonzero, which would allow model parameters to be identified.

## 6 Cost function minimization in identification

In this section we will show how identifiability properties of the model structure appear in gradient-based iterative parameter estimation algorithms. If we iteratively solve for a parameter estimate  $\hat{\theta}$  by minimizing a cost function  $V(\theta)$ , the general update rule in step  $m$  of a Newton-type algorithm is given by

$$\hat{\theta}_{m+1} = \hat{\theta}_m - \gamma \left( \frac{\partial^2 V}{\partial \theta^2} \right)^{-1} \frac{\partial V}{\partial \theta}, \quad (18)$$

where  $\gamma$  denotes a scalar damping factor. Note that in this expression the partial derivatives are evaluated in the local parameter  $\hat{\theta}_m$ . In contrast with the analysis in the previous section this local parameter does not necessarily reflect a (local) minimum of the cost function  $V$ .

If we consider the prediction error cost function as used before, then for the model structure considered and after linearization of  $\hat{\mathbf{y}}(\theta)$  around parameter  $\theta_m$  the update rule becomes

$$\hat{\theta}_{m+1} = \hat{\theta}_m + \gamma \left( \frac{\partial \hat{\mathbf{y}}(\theta)^T}{\partial \theta} \left( \frac{\partial \hat{\mathbf{y}}(\theta)^T}{\partial \theta} \right)^T \right)^{-1} \frac{\partial \hat{\mathbf{y}}(\theta)^T}{\partial \theta} (\mathbf{y} - \hat{\mathbf{y}}(\theta)). \quad (19)$$

where  $P_\gamma$  is considered identity for notational simplicity.

The parameter update (19) is actually a Gauss-Newton step ([5]), employing a first order Taylor expansion of  $\hat{\mathbf{y}}(\theta)$  around  $\theta_m$ , similar to the approximation in section 3.3. As an alternative, a Steepest-Descent algorithm ([5]) approximates the Hessian with any positive definite matrix, where standard the identity matrix is chosen. As a result, the update rule in the considered situation becomes:

$$\hat{\theta}_{m+1} = \hat{\theta}_m + \gamma \frac{\partial \hat{\mathbf{y}}(\theta)^T}{\partial \theta} (\mathbf{y} - \hat{\mathbf{y}}(\theta)).$$

If the model structure is not identifiable in  $\hat{\theta}_m$  the matrix inverse in (19) will not exist. Although this is often indicated as a serious problem for iterative optimization algorithms it can simply be overcome by restricting the update rule to make steps only in that part of the parameter space that does influence the output predictor, see e.g. [18, 28, 19]. This actually come down to utilizing the pseudo-inverse of the Jacobian in (19), on the basis of the SVD:

$$\frac{\partial \hat{\mathbf{y}}(\theta)^T}{\partial \theta} = [ U_1 \ U_2 ] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \quad (20)$$

with  $\Sigma_1 \in \mathbb{R}^{p \times p}$ . If  $\Sigma_2 = 0$ , the update rule for the Gauss-Newton iteration can then be replaced by

$$\hat{\theta}_{m+1} = \hat{\theta}_m + \gamma U_1 \Sigma_1^{-1} V_1^T (\mathbf{y} - \hat{\mathbf{y}}(\theta)),$$

while the update rule of Steepest-Descent is given by

$$\hat{\theta}_{m+1} = \hat{\theta}_m + \gamma U_1 \Sigma_1 V_1^T (\mathbf{y} - \hat{\mathbf{y}}(\theta)).$$

Both algorithms update the parameter only in the subspace that is determined by the column space of  $U_1$ , being the locally identifiable subspace of the parameter space in the local point  $\hat{\theta}_m$ .

Note that the difference between the two update mechanisms is that Steepest-Descent emphasizes the vectors of  $U_1$  that correspond to large singular values of the Jacobian, while Gauss-Newton emphasizes the vectors of  $U_1$  that correspond to small singular values of the Jacobian.

Large singular values of the Jacobian are associated with directions in which the predictions are very sensitive to a change in the parameters. Indeed, Steepest-Descent looks for the direction in which the cost function decreases with the least amount of effort in changing the parameters. Gauss-Newton follows an opposite strategy. Here the algorithm looks for a change in predicted outputs (i.e. cost function) which is induced by the largest change in the parameters.

It can simply be verified that parameter scaling does influence the estimate of the steepest-descent algorithm, in contrast with the Gauss-Newton algorithm which is scaling invariant, see e.g. [5]. This scaling-invariance however is only true if the Hessian has full rank in  $\theta = \theta_m$ .

Similar to the analysis in the previous sections the rank reduction of the Jacobian, as represented in (20) can of course be enforced if the SVD shows a large number of small singular values in  $\Sigma_2$ , and the Jacobian is approximated by setting  $\Sigma_2 = 0$ .

A similar approach of Jacobian reduction is employed in the fully parametrized state-space model identification using so-called data-driven local coordinates of [18, 19] as well as in subspace identification [28], where search directions are chosen to be orthogonal to the tangent space of the manifold representing equivalent models. See also [31] for a further comparison of methods. If the main interest of the modelling procedure is to identify (linear) system dynamics, these approaches are attractive as they simply use the parameters as vehicles to arrive at an appropriate system model. However, in this paper we aim at preserving the physical interpretation of the parameters and therefore are more focussing on the uniqueness of the parameters estimates in order to obtain reliable long-term (non-linear) model predictions.

## 7 A Bayesian approach

Lack of identifiability of a model structure and the subsequent non-uniqueness of parameters that are estimated on the basis of measurement data, can be dealt with in different ways. One way is to reduce the parameter space in the model structure, as indicated in the previous sections. Alternatively additional prior information can be added to the identification problem. In those situations where a parameter estimate may not be uniquely identifiable from the data, a regularization term can be added to the cost function that takes account of prior knowledge of the parameters to be estimated. In this setting an alternative (Bayesian) cost function is considered:

$$V_p(\theta) := V(\theta) + \frac{1}{2}(\theta - \theta_p)P_{\theta_p}^{-1}(\theta - \theta_p), \quad (21)$$

where the second term represents the weighted mismatch between the parameter vector and the prior parameter vector  $\theta_p$  with covariance  $P_{\theta_p}$ . When again the model output  $\hat{\mathbf{y}}(\theta)$  is approximated using a first-order Taylor expansion around  $\theta_p$ , the Hessian of (21) becomes:

$$\frac{\partial^2 V_p(\theta)}{\partial \theta^2} = \frac{\partial \hat{\mathbf{y}}(\theta)^T}{\partial \theta} P_v^{-1} \left( \frac{\partial \hat{\mathbf{y}}(\theta)^T}{\partial \theta} \right)^T + P_{\theta_p}^{-1}. \quad (22)$$

Since  $P_{\theta_p}^{-1}$  is positive definite by construction and the first term is positive semi-definite, the Hessian has full rank and the parameter estimate

$$\hat{\theta}_{bayes} = \arg \min_{\theta} V_p(\theta)$$

is unique. This uniqueness is guaranteed by the prior information that has been added to the problem. Formally there can still be lack of identifiability, however it is not any more reflected in a non-unique parameter estimate. A consequence of this approach is that the obtained parameter estimate may be highly influenced by the prior information, and less by the measurement data.

The covariance matrix of the Bayesian parameter estimate can also be analyzed using the classical prediction error theory, see [15]. Under ideal circumstances (consistent estimation and  $\theta_p = \theta_0$  (!)) it can be shown that

$$\text{cov}(\hat{\theta}_{bayes}) = \left[ \mathbb{E} \left. \frac{\partial^2 V_p(\theta)}{\partial \theta^2} \right|_{\theta_0} \right]^{-1}. \quad (23)$$

In other words, the inverse of the Hessian of the identification criterion remains to play the role of (sample estimate of) the parameter covariance matrix, and the same considerations as discussed in the earlier sections can be applied to the SVD analysis of this Hessian. By appropriately operating on the expression for the Hessian (22), it can be shown that a relevant SVD analysis for dimension reduction can now be applied to

$$P_{\theta_p}^T \frac{\partial \hat{\mathbf{y}}(\theta)^T}{\partial \theta} P_v^{-\frac{1}{2}},$$

which in [23] is referred to as the dimensionless sensitivity matrix.

It may be clear that the parameter estimate becomes highly dependent on the prior information, and that bias will occur when the parameter prior  $\theta_p$  is not correct.

It has to be noted that this Bayesian approach is typically followed when using sequential estimation algorithms for joint parameter and state estimation, as in Extended Kalman Filters and variations thereof, such as the Ensemble Kalman Filter, see e.g. [8].

## 8 Structural identifiability

The question whether parameters can be uniquely identified from data basically consists of two parts. The first part concerns the model structure: is it possible at all to distinguish two given parameters, provided that the input is chosen in the best possible way? This property is called structural identifiability of a model structure. The second part concerns the issue whether the actual input is informative enough to allow this distinction. In the previous sections both parts were considered simultaneously. In this section only the first part is investigated. Consider definition 2 on structural identifiability. Without loss of generality, but for ease of notations, we will limit attention to the SISO case. The multivariable case is also treated in [27]. Note that  $G(z, \theta)$  can be written as:

$$G(z, \theta) = \sum_{k=1}^{\infty} M(k, \theta) z^{-k}, \quad (24)$$

where  $M(k, \theta)$  are the Markov parameters. Based on (24) we argue that equality of  $G(z, \theta_1)$  and  $G(z, \theta_2)$  is related to equality of the Markov parameters of  $G(z, \theta_1)$  and  $G(z, \theta_2)$ . We can now use the following proposition ([9], [11], [20], [27]):

**Proposition 3** Consider the map  $S_N(\theta) : \Theta \subset \mathbb{R}^q \rightarrow \mathbb{R}^N$  defined by:

$$S_N(\theta) := [ M(1, \theta) \quad \dots \quad M(N, \theta) ]^T. \quad (25)$$

Then the model structure is locally structural identifiable in  $\theta_m$  if  $\text{rank} \left( \frac{\partial S_N^T(\theta)}{\partial \theta} \right) = q$  in  $\theta = \theta_m$ .

Both the qualitative question of structural identifiability, and the determination of the “best” structurally identifiable subspace of parameters can now be examined by applying an SVD to the matrix

$$\frac{\partial S_N^T(\theta)}{\partial \theta} \quad (26)$$

and examining the column space of this matrix, see [27]. However also in this problem we need to take care that our (approximate) identifiability test is not dependent on user-chosen parameter scaling, and so we need a premultiplication of (26) with the scaling matrix  $\Gamma_{\theta_m}$ . If a parameter has high impact on a particular Markov parameter, but the Markov parameter itself has a very small value, the considered parameter is still a good candidate to be removed in our model structure approximation problem. Therefore an additional weighting of (26) is desired that takes account of the values of the Markov parameters. As a result we consider the column space of the matrix

$$\Gamma_{\theta_m} \frac{\partial S_N^T(\theta)}{\partial \theta} \Gamma_S \quad (27)$$

where for the SISO case  $\Gamma_S := \text{diag} ( |M_1| \quad \dots \quad |M_N| )$ . The consequence is that Markov parameters that have a high value are considered to be more important to

include than Markov parameters with a small value.<sup>3</sup>

The row space of (27) that relates to the dominant singular values of the matrix, now is a representation of the parameter space of the approximated model structure. The Jacobian matrix (26) can be calculated analytically, as is shown in [27] for state space model structures.

The structurally identifiable problem and the identifiability problem are of course closely related to each other. This can be observed by realizing that

$$\frac{\partial \hat{\mathbf{y}}(\boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} = \frac{\partial S_N^T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \boldsymbol{\Phi}_N, \quad (28)$$

where  $\boldsymbol{\Phi}_N$  is given by

$$\boldsymbol{\Phi}_N = \begin{bmatrix} u_1 & u_2 & \dots & u_N \\ & u_1 & & \vdots \\ & & \ddots & \\ & & & u_1 \end{bmatrix} \quad (29)$$

and the derivatives are evaluated at  $\boldsymbol{\theta} = \boldsymbol{\theta}_m$ .

Note that the matrix  $\boldsymbol{\Phi}_N$  with input signals acts as a weighting matrix in (28) in a similar way as the weighting matrix  $\boldsymbol{I}_S$  does in (27).

## 9 Examples

In order to illustrate the concepts, and in particular the role of the scaling/weighting functions, we will now discuss two examples where we have chosen a very simple SISO finite impulse response (FIR) model. The model structure will be approximated using the previously discussed identifiability analysis, where we assume that  $P_v = I$ .

**Example 4** Consider the data-generating system

$$y(t) = \alpha_0 u(t-1) + \beta_0 u(t-2)$$

with  $\alpha_0 = 10^6$  and  $\beta_0 = 10^{-6}$ , and  $\boldsymbol{\theta}_0 := [\alpha_0 \ \beta_0]^T$ . Consider the input/output model structure

$$y(t, \boldsymbol{\theta}) = \alpha u(t-1) + \beta u(t-2), \quad \boldsymbol{\theta} := [\alpha \ \beta]^T.$$

For an analysis of the local identifiability in  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$  we consider

$$\boldsymbol{\psi}(t, \boldsymbol{\theta}_0) := \frac{\partial y(t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} u(t-1) \\ u(t-2) \end{bmatrix} \quad (30)$$

so that the Fisher information matrix  $\boldsymbol{J}$  is given by

---

<sup>3</sup> Note that in a more generalized setting this weighting should be replaced by a weighting that takes account of the application in which the model is used.

$$\begin{aligned} J &= \mathbb{E} \sum_{t=1}^N \psi(t, \theta_0) \psi(t, \theta_0)^T \\ &= N \mathbb{E} \begin{pmatrix} u(t-1) \\ u(t-2) \end{pmatrix} \begin{pmatrix} u(t-1) \\ u(t-2) \end{pmatrix}^T = N \begin{bmatrix} R_u(0) & R_u(1) \\ R_u(1) & R_u(0) \end{bmatrix} \end{aligned}$$

with  $R_u(\tau) = \mathbb{E}[u(t)u(t-\tau)]$ .

The scaled Fisher information matrix  $\tilde{J}$  of (14) for a local analysis around  $\theta_0$  is then given by

$$N \begin{bmatrix} \alpha_0 & 0 \\ 0 & \beta_0 \end{bmatrix} \begin{bmatrix} R_u(0) & R_u(1) \\ R_u(1) & R_u(0) \end{bmatrix} \begin{bmatrix} \alpha_0 & 0 \\ 0 & \beta_0 \end{bmatrix}.$$

The relative parameter variance is indicated by  $\tilde{J}^{-1}$ . In the case of a unit variance white noise input, it follows that

$$\tilde{J} = N \begin{bmatrix} 10^{12} & 0 \\ 0 & 10^{-12} \end{bmatrix},$$

while the unscaled Fisher information matrix satisfies  $J = N \cdot I$ . Analysis of  $\tilde{J}$  shows that the second parameter can very well be neglected, leading to an approximate model structure  $y(t) = \alpha u(t-1)$ .

Structural identifiability analysis without scaling shows that both parameters are structurally identifiable, since  $\frac{\partial S_N^T(\theta)}{\partial \theta} = I$ . However, including both scaling matrices  $\Gamma_\theta$  and  $\Gamma_S$ , we obtain

$$\Gamma_\theta \frac{\partial S_N^T(\theta)}{\partial \theta} \Gamma_S = \begin{bmatrix} 10^{12} & 0 \\ 0 & 10^{-12} \end{bmatrix},$$

also showing that the second parameter can be very well neglected. In light of section 5 we remark that in  $\theta = \theta_0$  this model is poorly observable/controllable and as a result it is also poorly identifiable.

**Example 5** In this example the same data-generating system as in the previous example is considered. Consider the input/output model structure

$$y(t, \theta) = \alpha u(t-1) + 10^{-6} \gamma u(t-2), \quad \theta := [\alpha \ \gamma]^T.$$

where  $\gamma_0 = 1$ . In comparison with the previous example we have scaled the second parameter with a factor  $10^{-6}$ . This can be thought of to be the result of choosing a different physical unit for the parameter. The scaled Fisher information matrix  $\tilde{J}$  of (14) is

$$N \begin{bmatrix} \alpha_0 & 0 \\ 0 & \gamma_0 \end{bmatrix} \begin{bmatrix} R_u(0) & 10^{-6} R_u(1) \\ 10^{-6} R_u(1) & 10^{-12} R_u(0) \end{bmatrix} \begin{bmatrix} \alpha_0 & 0 \\ 0 & \gamma_0 \end{bmatrix}.$$

Under the same input conditions it follows that

$$\tilde{J} = N \begin{bmatrix} 10^{12} & 0 \\ 0 & 10^{-12} \end{bmatrix},$$

while the unscaled Fisher information matrix is

$$J = N \begin{bmatrix} 1 & 0 \\ 0 & 10^{-12} \end{bmatrix}.$$

Whereas the unscaled matrix is essentially different from the previous example, the scaled analysis shows again that the second parameter can very well be neglected and that the model structure can be approximated with  $y(t) = \alpha u(t-1)$ .

Structural identifiability analysis without scaling shows that  $\alpha$  is structurally best identifiable, since

$$\frac{\partial S_N^T(\theta)}{\partial \theta} = \begin{bmatrix} 1 & 0 \\ 0 & 10^{-6} \end{bmatrix}.$$

Including both scaling matrices  $\Gamma_\theta$  and  $\Gamma_S$ , we obtain in quadratic form

$$\Gamma_\theta \frac{\partial S_N^T(\theta)}{\partial \theta} \Gamma_S = \begin{bmatrix} 10^{12} & 0 \\ 0 & 10^{-12} \end{bmatrix},$$

being exactly the same as matrix as in the previous example, meaning that the structural identifiability analysis is now-scaling invariant.

The examples are of course very simple, and they are meant to illustrate the basic phenomena that might occur in large scale physical structures. Use of a notion of relative variance, reflected in a scaled Fisher information matrix, leads to selection results that are scaling-invariant. The fact that we consider a local analysis only is of course a limitation of the results presented here.

## 10 Conclusions

The question whether a large scale (nonlinear) physical model structure is identifiable, is usually considered in a qualitative way. In this chapter the notion of identifiability is quantified and it is shown how the model structure can be approximated so as to achieve identifiability, while retaining the interpretation of the physical parameters. In this chapter this question has been addressed in a prediction error setting. The analysis has been related to iterative optimization algorithms (like Gauss-Newton and Steepest-Descent) and to Bayesian estimation. It has been shown that parameter scaling becomes relevant when approximating model structures.

**Acknowledgements** This chapter is dedicated to Okko Bosgra, a great colleague and supervisor, at the occasion of his 65th birthday, and with great appreciation for the many stimulating discussions we have had over the years.

## References

- [1] Bai, E.W.: An optimal two-stage identification algorithm for hammerstein-wiener nonlinear systems. *Automatica* **34**(3), 333–338 (1998)
- [2] Bamieh, B., Giarré, L.: Identification of linear parameter varying models. *Int. J. Robust and Nonlinear Control* **12**, 841–853 (2002)
- [3] Bellman, R., Åström, K.J.: On structural identifiability. *Mathematical Biosciences* **7**, 329–339 (1970)
- [4] Berntsen, H.E., Balchen, J.G.: Identifiability of linear dynamical systems. In: P. Eykhoff (ed.) *Proc. 3rd IFAC Symp. Identification and System Parameter Estimation*, pp. 871–874. The Hague (1973)
- [5] Dennis, J., Schnabel, R.: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, New Jersey (1983)
- [6] Dötsch, H.G.M., Van den Hof, P.M.J.: Test for local structural identifiability of high-order non-linearly parameterized state space models. *Automatica* **32**(6), 875–883 (1996)
- [7] Evans, N., Chapman, M., Chappell, M., Godfrey, K.: Identifiability of uncorrelated nonlinear rational systems. *Automatica* **38**, 1799–1805 (2002)
- [8] Evensen, G.: *Data Assimilation*. Springer (2007)
- [9] Glover, K., Willems, J.C.: Parameterizations of linear dynamical systems: canonical forms and identifiability. *IEEE Trans. Automatic Control* **19**(6), 640–646 (1974)
- [10] Godfrey, K.R.: *Compartmental Models and their Application*. Academic Press, London (1983)
- [11] Grewal, M.S., Glover, K.: Identifiability of linear and nonlinear dynamical systems. *IEEE Trans. Automatic Control* **21**(6), 833–837 (1976)
- [12] Hjalmarsson, H.: From experiment design to closed-loop control. *Automatica* **41**(3), 393–438 (2005)
- [13] Jansen, J.D., Bosgra, O.H., Van den Hof, P.M.J.: Model-based control of multiphase flow in subsurface oil reservoirs. *J. Process Control* **18**, 846–855 (2008)
- [14] Lee, L.H., Poolla, K.: Identification of linear parameter-varying systems using nonlinear programming. *J. of Dynamic Systems, Measurement and Control* **121**(1), 71–78 (1999)
- [15] Ljung, L.: *System Identification: Theory for the User*, 2nd edition. Prentice-Hall, Englewood Cliffs, NJ (1999)
- [16] Ljung, L., Glad, T.: On global identifiability for arbitrary model parameterizations. *Automatica* **30**(2), 265–276 (1994)
- [17] Lund, B.F., Foss, B.A.: Parameter ranking by orthogonalization - applied to nonlinear mechanistic models. *Automatica* **44**, 278–281 (2008)
- [18] McKelvey, T., Helmersson, A.: System parameterization of an overparameterized model class: improving the optimization algorithm. In: *36th IEEE Conf. Decision and Control*, pp. 2984–2989. San Diego, CA, USA (1997)
- [19] McKelvey, T., Helmersson, A., Ribarits, T.: Data driven local coordinates for multivariable linear systems and their application to system identification. *Automatica* **40**, 1629–1635 (2004)
- [20] Norton, J.P.: Normal-mode identifiability analysis of linear compartmental systems in linear stages. *Mathematical Biosciences* **50**, 95–115 (1980)
- [21] Pintelon, R., Schoukens, J.: *System Identification: A Frequency Domain Approach*. IEEE Press, Piscataway, NJ (2001)
- [22] Stigter, J.D., Peeters, R.L.M.: On a geometric approach to the structural identifiability problem and its application in a water quality case study. In: *Proc. European Control Conf. Koss, Greece* (2007)
- [23] Tavakoli, R., Reynolds, A.C.: History matching with parameterization based on the SVD of a dimensionless sensitivity matrix. In: *Proc. SPE Reservoir Simulation Symposium, Texas, USA* (2009)
- [24] Tóth, R.: *Modeling and identification of linear parameter-varying systems - an orthonormal basis function approach*. Ph.D. thesis, Delft Univ. Technology, Delft, The Netherlands (2008)
- [25] Tóth, R., Heuberger, P.S.C., Van den Hof, P.M.J.: LPV system identification with globally fixed orthonormal basis functions. In: *Proc. 46th IEEE Conf. Decision and Control*, pp. 3646–3653. New Orleans, LA (2007)

- [26] Vajda, S., Rabitz, H., Walter, E., Lecourtier, Y.: Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. *Chem. Eng. Comm.* **83**, 191–219 (1989)
- [27] Van Doren, J.F.M., Van den Hof, P.M.J., Jansen, J.D., Bosgra, O.H.: Determining identifiable parameterizations for large-scale physical models in reservoir engineering. In: M. Chung, P. Misra, H. Shim (eds.) *Proc. 17th IFAC World Congress*, pp. 11,421–11,426. Seoul (2008)
- [28] Verdult, V.: Nonlinear system identification: A state-space approach. Ph.D. thesis, Delft Univ. Technology, Delft, The Netherlands (2002)
- [29] Verdult, V., Verhaegen, M.: Subspace identification of multivariable linear parameter-varying systems. *Automatica* **38**(5), 805–814 (2002)
- [30] Walter, E.: *Identifiability of Parametric Models*. Pergamon Press, Oxford (1987)
- [31] Wills, A., Ninness, B.M.: On gradient-based search for multivariable system estimates. *IEEE Trans. Automatic Control* **53**(1), 298–306 (2008)
- [32] van Wingerden, J., Verhaegen, M.: Subspace identification of bilinear and LPV systems for open and closed loop data. *Automatica* **45** (2009). To appear
- [33] Zhu, Y.C.: Estimation of an N-L-N Hammerstein-Wiener model. *Automatica* **38**(9), 1607–1614 (2002)
- [34] Zhu, Y.C., Xu, X.: A method of LPV model identification for control. In: M. Chung, P. Misra, H. Shim (eds.) *Proc. 17th IFAC World Congress*, pp. 5018–5023. Seoul (2008)