9

Nonlinear System Identification through Local Model Approaches: Partitioning Strategies and Parameter Estimation

Christoph Hametner and Stefan Jakubek
Vienna University of Technology
Austria

1. Introduction

In this chapter nonlinear system identification utilising Total Least Squares (TLS) and Generalised Total Least Squares (GTLS) methodologies in local model networks is addressed. These Neuro-Fuzzy networks are based on the identification of subdomains of the system that can reasonably accurately be described by local models. The aggregation of such subdomains in a so-called local model network then yields a versatile description of the overall system. One of the main challenges in the design of local model networks is the determination of the region of validity for the local models. Some of the related methodologies completely leave the partitioning to the user so that a solid idea about the nature of the nonlinearity of the system is required, cf. Johansen et al. (2000). Other methods make use of the input/output data of the system to identify suitable subdomains, Jakubek & Keuth (2006); Nelles (2002).
The identification and partitioning algorithm presented in this contribution is based on an iterative decomposition that works in the partition space rather than in the input- or product space. Thereby, in each step an axis-oblique partitioning is performed by multi-objective optimisation using an Expectation-Maximisation (EM) algorithm, Hametner & Jakubek (2007). The discrimination between input space and partition space allows the incorporation of prior knowledge into the partitioning process and reduces the complexity of the optimisation problem dramatically. In contrast to conventional clustering algorithms the number of rules is determined during the training by adding new models to the hierarchical tree until no statistically significant improvement is achieved.
The second important problem addressed in this work is the estimation of the local model parameters in the presence of noise in measured input *and* output data. In this situation conventional parameter estimation methods generally do not yield consistent estimates. Methods that are designed to cope with the case of noisy input and output channels are generally called "errors-in-variables methods", Soderstrom (2007). The corresponding block diagram is depicted in Fig. 1.
Various approaches for the identification of linear systems in a complex noisy environment can be found in literature, e. g. Han et al. (1996); Vandersteen (1998). A well-suited estimation procedure when all inputs are subject to noise is the Total Least Squares method (TLS). This method has been studied in many different areas over the last 25 years, e. g. Golub & Loan (1980); Huffel & Zha (1991); Markovsky et al. (2005). The main drawback of TLS is that

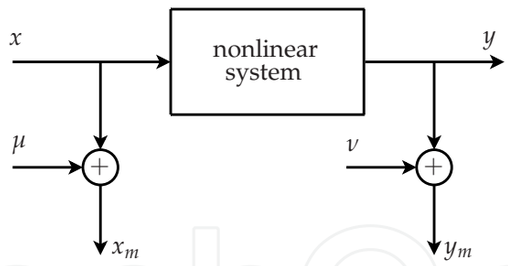


Fig. 1. Errors-in-variables model

strongly different noise levels easily result in ill-posed problems. A proper extension to the TLS method for such cases are Generalized Total Least Squares methods (GTLS). Generally, GTLS algorithms are methods for *linear* parameter estimation, when some or all inputs are subject to noise. There are various analyses and solution approaches to solve the GTLS problem, Huffel & Zha (1991); Nayak et al. (2006); Paige & Wei (1993); Van Huffel & J. (1989). For the integration of the above mentioned methodologies in a local model network weighted TLS and GTLS parameter estimation algorithms are presented in this work that allow for individual weighting of data records.

Recent local model network approaches utilise the prediction error for partitioning, Abonyi et al. (2002); Hametner & Jakubek (2007); Jakubek & Keuth (2006). If some or all signals involved in the parameter estimation process are corrupted by noise this approach is no longer feasible. In this paper a more general residual is defined to determine the region of validity of the local models. As a basis for the partitioning procedure mentioned above a suitable formulation of the *GTLS residual* will be introduced.

This chapter is organised as follows: Section 2 describes the architecture and the construction of the local model network. In section 3 weighted TLS and GTLS parameter estimation algorithms and the associated residual are presented. In the last section 4 the applicability and benefits of the proposed concepts are demonstrated by means of an illustrative example.

2. Construction of the local model network

Local model networks offer a versatile structure for the identification of nonlinear static and dynamic systems. The construction of these Neuro-Fuzzy networks is based on the identification of subdomains of the system that can reasonably accurate be described by local models. The aggregation of such subdomains then yields the description of the overall system.

The architecture of a dynamic local model networks is depicted in Fig. 2. The *physical inputs* are denoted by $\boldsymbol{u} = \begin{bmatrix} u_1 & u_2 & \dots & u_q \end{bmatrix}^T$ and the output by y, respectively. Each local model (indicated by subscript i) consists of two parts: The validity function Φ_i and its model parameters $\boldsymbol{\theta}_i$.

The *local* estimate for the output $y(k)$ is obtained by

$$\hat{y}_i(k) = \boldsymbol{x}^T(k)\boldsymbol{\theta}_i, \tag{1}$$

where $\boldsymbol{x}(k)$ denotes the input vector for the rule consequents at time k. For *dynamic* models typically a local affine model structure is implemented and $\boldsymbol{x}(k)$ contains past inputs and outputs and the offset term:

$$\boldsymbol{x}(k) = \begin{bmatrix} u_1(k-1) \\ \cdots \\ u_q(k-m) \\ y(k-1) \\ \cdots \\ y(k-n) \\ 1 \end{bmatrix} . \tag{2}$$

In (2) m and n denote the system order of the numerator and denominator respectively. For ease of demonstration the time delay of the system is neglected in (2). In case of *static* modelling the regressor vector $\boldsymbol{x}(k)$ may contain arbitrary functions (e. g. polynomials) of the elements of $\boldsymbol{u}$.

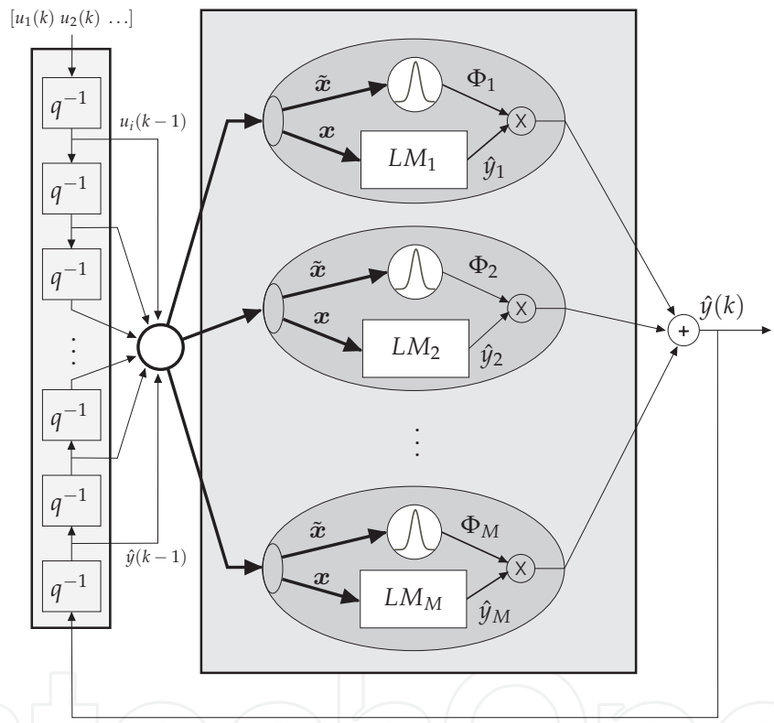


Fig. 2. Architecture of a dynamic local model network

All local estimations $\hat{y}_i(k)$ are used to form the global model output $\hat{y}(k)$ by weighted aggregation

$$\hat{y}(k) = \sum_{i=1}^{M} \Phi_i(k)\hat{y}_i(k), \tag{3}$$

where

$$\Phi_i(k) = \Phi_i(\tilde{\boldsymbol{x}}(k)) \quad , \quad \hat{y}_i(k) = \boldsymbol{x}^T(k)\boldsymbol{\theta}_i \tag{4}$$

and M denotes the number of local models.
Obviously, the input vector for the membership functions $\tilde{\boldsymbol{x}}(k)$ can be chosen differently to the input vector for the local model, cf. Fig. 2. The discrimination between input arguments of the

consequents and the premises is particularly useful if information about the structure of the nonlinearity is available, Nelles (2002). Especially in dynamic identification the dimension of the partition space and furthermore the complexity of the optimisation problem can be reduced dramatically. Therefore, the elements of the partition space $\tilde{\boldsymbol{x}}(k)$ are chosen according to the nonlinear behaviour of the system only.
The algorithm outlined here is based on the concept of decision trees, Breiman et al. (1984); Theodoridis & Koutroumbas (1999). The growing tree can be described by a binary tree where each node corresponds to a split of the partition space into two parts, e. g. Fig. 3. The free ends of the branches represent the actual local models with their validity functions Φ_i and the parameter vectors $\boldsymbol{\theta}_i$. The validity functions for the layout in Fig. 3 are obtained by

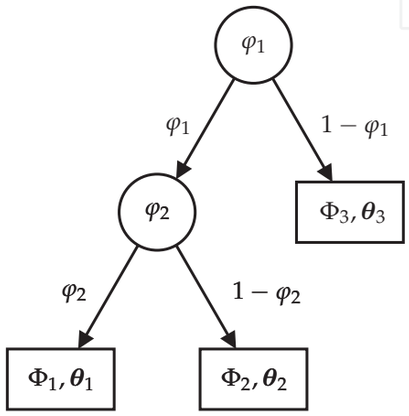


Fig. 3. Hierarchical discriminant tree

$$\Phi_1(\tilde{\boldsymbol{x}}) = \varphi_1(\tilde{\boldsymbol{x}})\varphi_2(\tilde{\boldsymbol{x}}), \tag{5}$$

$$\Phi_2(\tilde{\boldsymbol{x}}) = \varphi_1(\tilde{\boldsymbol{x}})(1-\varphi_2(\tilde{\boldsymbol{x}})), \tag{6}$$

$$\Phi_3(\tilde{\boldsymbol{x}}) = 1-\varphi_1(\tilde{\boldsymbol{x}}). \tag{7}$$

The partitioning of these new models is that a partition of unity ($\sum_{i=1}^{M} \Phi_i(\tilde{\boldsymbol{x}}) = 1$) throughout the partition space (for all $\tilde{\boldsymbol{x}}$) is guaranteed and no normalisation side effects like reactivations can occur cf. Nelles (2002).
The main challenge with local model networks is to determine Φ_i in such a way that the local estimate (1) sufficiently accurate describes the true process within the region of validity.
In each iteration step one local model of the growing tree is replaced by a new node and two new models attached to this node. The selection of the particular local model to be replaced is based on local accuracy which is assessed by the mean squared GTLS residual (see section 3.3). The partitioning of this new models is obtained considering a two-category classification problem. The main advantage of this concept is that with each iteration step the number of training data involved and thus the computational effort decreases.

2.1 Discriminant optimisation

In each step of the recursive decomposition of the partition space a nonlinear optimisation is necessary. Considering the two-category classification problem each data point has to be assigned to either class η_1 or η_2 depending on the value of the corresponding discriminants $\varphi_i(\tilde{\boldsymbol{x}}(k), \boldsymbol{\psi})$:

$$\varphi_1(\tilde{\boldsymbol{x}}(k), \boldsymbol{\psi}) = \frac{1}{1 + \exp(-\begin{bmatrix}1 & \tilde{\boldsymbol{x}}^T(k)\end{bmatrix} \boldsymbol{\psi})} \text{ for class } \eta_1 \tag{8}$$

$$\varphi_2(\tilde{\boldsymbol{x}}(k), \boldsymbol{\psi}) = 1 - \varphi_1(\tilde{\boldsymbol{x}}(k), \boldsymbol{\psi}) \text{ for class } \eta_2. \tag{9}$$

Here, $\boldsymbol{\psi} = \begin{bmatrix}\psi_0 & \psi_1 & \dots & \psi_p\end{bmatrix}^T$ denotes the weight vector. Each node of the model tree (see Fig. 3) corresponds to a logistic sigmoid discriminant function which is depicted in Fig. 4.

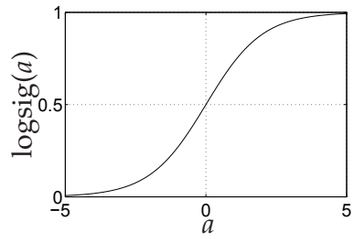


Fig. 4. Logistic sigmoid discriminant function

Since the input argument into (8) is linear in its input $\tilde{\boldsymbol{x}}(k)$ the decision boundary is also linear in the partition space, e. g. Fig. 5.

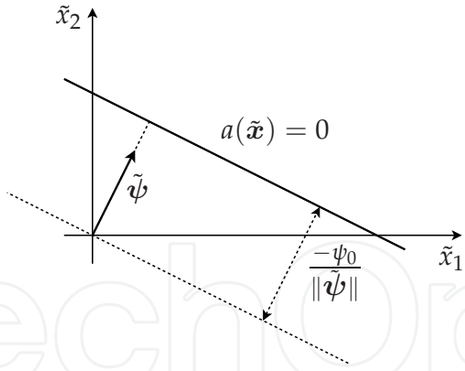


Fig. 5. Linear decision boundary of the nonlinear discriminant function

In order to find an optimal weight vector $\boldsymbol{\psi}$ for the two-category classification problem, an Expectation-Maximisation based algorithm is applied, e. g. Dempster et al. (1977). The goal of the iterative EM algorithm is to classify each data point in a way as to minimise the probability of misclassification. The EM algorithm proceeds in two steps (see also Hametner & Jakubek (2007)):

E-step: Based on the current estimate of the weight vector $\boldsymbol{\psi}_n$ the posterior probabilities $p(\eta_i|\boldsymbol{w}^T(k), \boldsymbol{\psi}_n)$ are calculated: The parameters of the two classes (local models) are

computed using weighted TLS or GTLS. For the evaluation of the model error a normal density function based on the (G)TLS residual, which is defined in section 3.3, is calculated. With the class-conditional residual distribution function (43), the posterior probabilities are obtained using Bayes' theorem.

M-step: The posterior probabilities are used to compute the new weight vector for the discriminant function and parameters of the two classes. The maximisation aims at determining the optimal weight parameters for the logistic discriminant function. Using e. g. a Levenberg-Marquardt learning algorithm the parameter vector $\boldsymbol{\psi}_n$ is adjusted such that the discriminant function (8) is optimally fit to $p(\eta_i|\boldsymbol{w}^T(k), \boldsymbol{\psi}_n)$.

Following, the steps of the iterative optimisation algorithm are reformulated:

2.1.1 Iterative optimisation algorithm

- *Initial choice* for $\boldsymbol{\psi}$: $\boldsymbol{\psi}_{ini}$ is ideally chosen such that the discriminant function bisects the data to be modeled. Thus the robustness to outliers is improved because the generation of two models with very unequal dimensions is avoided.
- *Step 1:* Compute the posterior probabilities $p(\eta_i|\boldsymbol{w}^T(k), \boldsymbol{\psi}_n)$.
- *Step 2:* Calculate/optimise the local model parameters $\boldsymbol{\theta}_{1,2}$ of the two classes by weighted (G)TLS.
- *Step 3:* Adjust $\boldsymbol{\psi}$ by Levenberg-Marquardt.
- *Repeat steps 1-3* until a certain termination criterion is reached, e. g. $\| \boldsymbol{\psi}_n - \boldsymbol{\psi}_{n-1}\|_2 < \epsilon$, where n denotes the iteration step and ϵ is a certain termination tolerance.

3. Local model parameter estimation

In many applications the inputs and outputs of a system are taken from measurements and are thus subject to noise. In this situation conventional parameter estimation methods suffer from the drawback that the parameter estimates are not consistent.
In this section weighted TLS and GTLS algorithms for the estimation of the local model parameters are presented. The application of Total Least Squares for parameter estimation from noisy inputs and outputs has been suggested repeatedly in recent years, Heij C. (1999); Markovsky et al. (2005); Roorda (1995). Demonstrative introductions and derivations of TLS are given in e. g. De Groen (1996); Golub & Loan (1980); Markovsky & Huffel (2007). While TLS is limited to situations where *all* channels are subject to noise, the GTLS algorithm yields consistent parameter estimates in the more general case when *some* (or all) channels are corrupted by noise, see also Huffel & Zha (1991); Nayak et al. (2006); Paige & Wei (1993); Van Huffel & J. (1989).
For the integration of the above mentioned methodologies in a local model network the algorithms presented in this section allow for individual weighting of data records.

3.1 Weighted Total Least Squares

The estimation of the local model parameters by Least Squares (LS) is based on the minimisation of the prediction error at the training data:

$$J = \frac{1}{2N} \sum_{k=1}^{N} [y(k) - \hat{y}(k)]^2. \tag{10}$$

Here N denotes the number of data records used for the local model.
In the case that only target data are affected by noise the minimisation of (10) yields a bias-free estimation of $\boldsymbol{\theta}_i$. Fig. 6 illustrates the problem by means of a simple linear map $y(x) = \beta_1 x + \beta_0$ where both x and y are corrupted by noise. For the parameter estimation only measured data $y_m = y + \nu$ and $x_m = x + \mu$ are available. For the ease of demonstration the noise variances $\sigma^2(\nu)$ and $\sigma^2(\mu)$ are assumed to be equal and uncorrelated. Fig. 6 compares the model obtained by minimisation of (10) to the true process.
In order to obtain a bias-free parameter estimation in the case of noisy inputs and outputs it is necessary to reconstruct both outputs *and* inputs. This means that instead of (10) the following criterion has to be minimised:

$$J = \frac{1}{2N} \left\{ \sum_{k=1}^{N} [x(k) - \hat{x}(k)]^2 + \sum_{k=1}^{N} [y(k) - \hat{y}(k)]^2 \right\} \tag{11}$$

Since (11) entails that both inputs and outputs have to be reconstructed the underlying optimisation is called *Total Least Squares* (TLS). From a geometric point of view the optimisation of (11) requires that the euclidean distances between data points and the model are minimised. Fig. 7 shows that this approach significantly improves the accuracy of the model. It can be shown that for $N \to \infty$ TLS delivers bias-free parameter estimates, cf. Heij C. (1999).

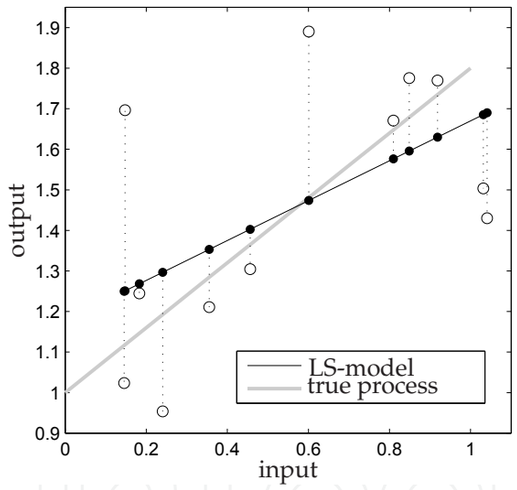


Fig. 6. Linear model, optimised by Least Squares

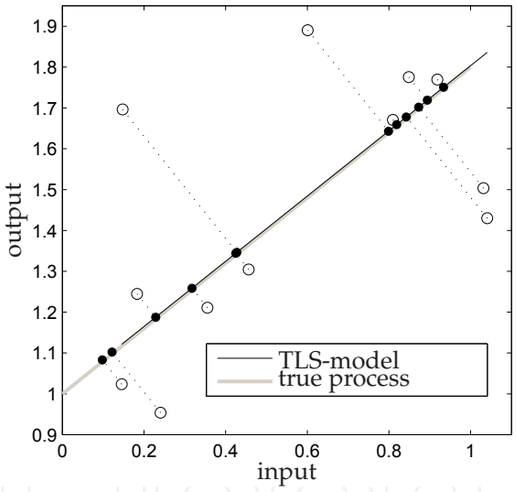


Fig. 7. Linear model, optimised by Total Least Squares

Following, an exemplary explanation of Total Least-Squares for a linear dynamic system is presented:
Let $\boldsymbol{X} \in \mathbb{R}^{N \times M}$ be the regressor matrix where every row contains all elements of $\boldsymbol{x}^T(k)$ except the constant offset term and let $\boldsymbol{y} \in \mathbb{R}^{N \times 1}$ be the observation vector. TLS aims at modifying both $\boldsymbol{y}$ and $\boldsymbol{X}$ in such a way that the following condition is satisfied:

$$\hat{\boldsymbol{y}} \in \text{Image}(\hat{\boldsymbol{X}}) \;\text{ and }\; ||\, \boldsymbol{y} - \hat{\boldsymbol{y}} \mid \boldsymbol{X} - \hat{\boldsymbol{X}} \,||_F = \min\,. \tag{12}$$

For the reconstructions $\hat{\boldsymbol{y}}, \hat{\boldsymbol{X}}$ a linear affine model structure is chosen. For that purpose an augmented regressor matrix is defined:

$$\boldsymbol{W} = [\, \boldsymbol{y} \mid \boldsymbol{X} \,] \tag{13}$$

A linear affine reconstruction of $\boldsymbol{w}^T$ is obtained from

$$\hat{\boldsymbol{w}}^T = \boldsymbol{w}^T - \left[(\boldsymbol{w}-\boldsymbol{m})^T\boldsymbol{b}\right]\boldsymbol{b}^T \;\text{ with } ||\boldsymbol{b}||_2 = 1 \tag{14}$$

Here $\boldsymbol{b}$ denotes the unit normal vector to the affine hyperplane and $\boldsymbol{m}$ is a point the hyperplane passes through. This linear model structure also ensures that $\hat{\boldsymbol{y}} \in \text{Image}(\hat{\boldsymbol{X}})$ holds. A reconstruction of all N data records yields

$$\hat{\boldsymbol{W}} = \boldsymbol{W} - \left[(\boldsymbol{W}-\mathbf{1}\boldsymbol{m}^T)\boldsymbol{b}\right]\boldsymbol{b}^T \tag{15}$$

with the $N \times 1$-vector $\mathbf{1} = [\,1\,,\,1\,,\,\ldots\,,\,1]^T$. The difference between original data and their reconstructions according to (12) is then given by

$$\boldsymbol{W} - \hat{\boldsymbol{W}} = \left[(\boldsymbol{W}-\mathbf{1}\boldsymbol{m}^T)\boldsymbol{b}\right]\boldsymbol{b}^T. \tag{16}$$

For *weighted* parameter estimation the estimation errors are weighted through the validity function Φ_i. Let $\boldsymbol{Q}_i$ denote a diagonal weighting matrix for the i-th local model. Its diagonal elements $q_i(k)$ represent the values of the validity function at the training data points. Compared to (11) a modified criterion is defined:

$$J_i = \frac{1}{2N}\left\{\sum_{k=1}^{N} q_i(k)[x(k)-\hat{x}(k)]^2 + \sum_{k=1}^{N} q_i(k)[y(k)-\hat{y}(k)]^2\right\} \tag{17}$$

Minimisation of (17) corresponds to a weighted version of TLS. Instead of the Frobenius norm in (12) one now has to minimise the following norm:

$$||\boldsymbol{Q}_i^{1/2}(\boldsymbol{W}-\hat{\boldsymbol{W}})||_F^2.$$

Since $||\boldsymbol{b}||_2 = 1$ the Frobenius norm becomes

$$||\boldsymbol{Q}_i^{1/2}(\boldsymbol{W}-\hat{\boldsymbol{W}})||_F^2 =$$

$$= \boldsymbol{b}^T(\boldsymbol{W}^T - \boldsymbol{m}\mathbf{1}^T)\boldsymbol{Q}_i(\boldsymbol{W}-\mathbf{1}\boldsymbol{m}^T)\boldsymbol{b}. \tag{18}$$

For the determination of $\boldsymbol{b}$ and $\boldsymbol{m}$ a weighted centroid vector $\boldsymbol{\mu}_W$ of all data records is defined:

$$\boldsymbol{\mu}_{Wi} = \boldsymbol{W}^T\boldsymbol{q}_i/s_q. \tag{19}$$

Here $\boldsymbol{q}_i$ denotes the vector composed from the main diagonal of $\boldsymbol{Q}_i$ and s_q is the sum of its elements: $s_q = \mathbf{1}^T\boldsymbol{q}_i$.

If all data records are referenced to the centroid (19) according to $\tilde{\boldsymbol{W}} = \boldsymbol{W} - \mathbf{1}\boldsymbol{\mu}_W^T$ and if one further observes that referencing to the centroid yields $\boldsymbol{q}_i^T\tilde{\boldsymbol{W}} = \mathbf{0}$ the minimisation of (18) yields

$$||\boldsymbol{Q}_i^{1/2}(\boldsymbol{W}-\hat{\boldsymbol{W}})||_F^2 = \boldsymbol{b}^T(\tilde{\boldsymbol{W}}^T\boldsymbol{Q}_i\tilde{\boldsymbol{W}})\boldsymbol{b} +$$

$$+ s_q[(\boldsymbol{m}-\boldsymbol{\mu}_{Wi})^T\boldsymbol{b}]^2 = \min. \tag{20}$$

The optimal choice for $\boldsymbol{m}$ is apparently $\boldsymbol{m} = \boldsymbol{\mu}_{Wi}$. The matrix $\tilde{\boldsymbol{W}}^T \boldsymbol{Q}_i \tilde{\boldsymbol{W}}$ is symmetric and positive semidefinite, the unit normal vector $\boldsymbol{b}$ is consequently obtained as the eigenvector associated to the smallest eigenvalue of $\tilde{\boldsymbol{W}}^T \boldsymbol{Q}_i \tilde{\boldsymbol{W}}$.
After partitioning of $\boldsymbol{b}$ according to $\boldsymbol{b} = [b_1\ ,\ \boldsymbol{\beta}]^T$ TLS offers the following difference equation:

$$\hat{y}_i(k) = \frac{1}{b_1}\left[-\boldsymbol{x}^T(k)\boldsymbol{\beta} + \boldsymbol{m}^T\boldsymbol{b}\right]. \tag{21}$$

3.1.1 Decorrelation of identification data

In the minimisation of (11) it was assumed that all measurements in $\boldsymbol{w}^T$ are equally corrupted with noise and that the individual noise sources are uncorrelated. In practical applications these prerequisites are almost never fulfilled which can lead to a misinterpretation by the TLS parameter estimation concept, cf. Abonyi et al. (2002). In these cases the identification data must be decorrelated prior to parameter identification.
Let $\nu(k)$ denote the noise signal that is superimposed to the true output $y_0(k)$ and let $\mu(k)$ be the noise signal belonging to the true input $u_0(k)$. The data record $\boldsymbol{w}(k)$ is the obtained from its unperturbed equivalent $\boldsymbol{w}_0(k)$ from

$$\boldsymbol{w}^T(k) = \boldsymbol{w}_0^T(k) + \boldsymbol{n}^T(k),$$

with the noise vector

$$\begin{aligned}\boldsymbol{n}(k) = [\,&\nu(k)\,,\,\nu(k-1)\,,\,\ldots\\ |\,&\mu(k-d-1)\,,\,\mu(k-d-2)\,,\,\ldots]^T.\end{aligned} \tag{22}$$

The covariance matrix $\boldsymbol{R}_n = E\{\boldsymbol{n}(k)\boldsymbol{n}^T(k)\}$ contains all the above mentioned correlations and is assumed to be known. In practical applications it can be determined from data records $\tilde{\boldsymbol{W}}_s$ from steady-state phases according to

$$\boldsymbol{R}_n \approx \frac{1}{N-1}(\tilde{\boldsymbol{W}}_s^T\tilde{\boldsymbol{W}}_s). \tag{23}$$

For $N \to \infty$ the approximation (23) converges to the expectation $E\{\boldsymbol{n}(k)\boldsymbol{n}^T(k)\}$. In the practical tests conducted in connection with the presented method it turned out that (23) always led to good results. For a correct application of TLS the following statistical property must hold:

$$\frac{1}{N-1}E\{(\tilde{\boldsymbol{W}} - \boldsymbol{W}_0)^T(\tilde{\boldsymbol{W}} - \boldsymbol{W}_0)\} = \boldsymbol{I},$$

where

$$\boldsymbol{W}_0 = \begin{bmatrix}\boldsymbol{w}_0^T(1)\\ \boldsymbol{w}_0^T(2)\\ \vdots\\ \boldsymbol{w}_0^T(N)\end{bmatrix}. \tag{24}$$

In order to accomplish this $\boldsymbol{b}$ is substituted by $\boldsymbol{b} = \boldsymbol{T}\tilde{\boldsymbol{b}}$. The new relevant noise covariance matrix then becomes

$$\tilde{\boldsymbol{R}} = \boldsymbol{T}^T\boldsymbol{R}_n\boldsymbol{T}. \tag{25}$$

A correct optimisation of $\tilde{\boldsymbol{b}}$ through TLS can thus be assured if the transformation matrix $\boldsymbol{T}$ is chosen such that

$$\boldsymbol{T}^T\boldsymbol{R}_n\boldsymbol{T} = \boldsymbol{I} \tag{26}$$

holds. The transformed unit normal vector $\tilde{\boldsymbol{b}}$ then turns out as the eigenvector belonging to the smallest eigenvalue of the matrix $\boldsymbol{T}^T\tilde{\boldsymbol{W}}^T\boldsymbol{Q}_i\tilde{\boldsymbol{W}}\boldsymbol{T}$.
In the case of pure measurement noise $\boldsymbol{R}_n$ is a diagonal matrix so that $\boldsymbol{T} = \boldsymbol{R}_n^{-1/2}$ is a solution to (26). Otherwise $\boldsymbol{T}$ can be obtained from the inverse of the Cholesky factorization of $\boldsymbol{R}_n$.

3.2 Weighted Generalised Total Least Squares

The GTLS algorithm described in this paper is based on the idea that conventional TLS inherently reconstructs all noisy inputs and the output.
As stated above the minimisation of (11) only leads to a consistent parameter estimate if all signals are corrupted by noise with equal variance and zero cross-correlations. If single variances are very different or if some signals are noise free the necessary decorrelation (see section 3.1.1) can easily result in an ill-posed problem. The GTLS algorithm overcomes these difficulties by excluding such signals from the reconstruction and treating them as noise-free. It is worth mentioning that the algorithm does not discriminate between inputs (regressors) and the output (observation). Consequently even the observation vector itself can be regarded as noise-free in the parameter estimation process.
As opposed to TLS, for GTLS parameter estimation noisy components in the augmented regressor $\boldsymbol{W}$ (see equation (13)) are indicated by a subscript "n" and noise-free components by an "o". Accordingly, $\boldsymbol{W}$ is partitioned into a part containing only noisy components $\boldsymbol{W}_n$ and another part containing only noise-free components $\boldsymbol{W}_o$:

$$\boldsymbol{W} = \begin{bmatrix}\boldsymbol{W}_n & \boldsymbol{W}_o\end{bmatrix}. \tag{27}$$

Without loss of generality, (27) is obtained from (13) by appropriately reordering its columns. In the sequel, $\boldsymbol{W}$ (and its estimates) always refer to the definition given by (27).
Similarly to TLS it is assumed that the noise signal $\boldsymbol{n}^T(k)$ that corrupts the rows of $\boldsymbol{W}_n$ is Gaussian with unity variance and zero cross-correlations. In most practical cases this has to be ensured by decorrelation of the noisy regressors, see section 3.1.1.
As opposed to TLS, GTLS only reconstructs noisy components:

$$\hat{\boldsymbol{W}} = \hat{\boldsymbol{W}}_n. \tag{28}$$

In the algorithm presented here, the reconstruction is based on ordinary TLS estimation where the original noisy data are projected on the TLS hyperplane using a reference point $\boldsymbol{m}$ and a unit normal vector $\boldsymbol{b}$, see equation (14). GTLS uses as an augmentation a linear combination of the noise-free regressors, given by $\boldsymbol{C}^T\boldsymbol{w}_o$. Here $\boldsymbol{w}_o^T$ denotes the noise-free regressor which is one row vector of $\boldsymbol{W}_o$ in (27).
The number of rows in $\boldsymbol{C}$ corresponds to the number of noise-free signals whereas the number of columns is the number of noisy signals.
Thus the GTLS reconstruction becomes

$$\hat{\boldsymbol{w}}_n^T(k) = \boldsymbol{w}_n^T(k) - \left[(\boldsymbol{w}_n(k) - \boldsymbol{m} - \boldsymbol{C}^T\boldsymbol{w}_o(k))^T\boldsymbol{b}\right]\boldsymbol{b}^T \tag{29}$$

and re-written for all data records ($k = 1, 2, \ldots N$)

$$\boldsymbol{W}_n - \hat{\boldsymbol{W}}_n = \left[(\boldsymbol{W}_n - \mathbf{1}\boldsymbol{m}^T - \boldsymbol{W}_o\boldsymbol{C})\boldsymbol{b}\right]\boldsymbol{b}^T, \tag{30}$$

with the $(N \times 1)$-vector $\mathbf{1} = \begin{bmatrix}1 & 1 & \ldots & 1\end{bmatrix}^T$. Optimisation of parameters yields

$$\|\boldsymbol{W}_n - \hat{\boldsymbol{W}}_n\|_F^2 = \min \quad \text{subject to} \quad \hat{\boldsymbol{y}} \in \text{Image}(\hat{\boldsymbol{X}}). \tag{31}$$

According to (31) the *weighted* optimisation of parameters yields

$$\|\boldsymbol{Q}_i^{1/2}(\boldsymbol{W}_n - \hat{\boldsymbol{W}}_n)\|_F^2 = \min \quad \text{subject to} \quad \hat{\boldsymbol{y}} \in \text{Image}(\hat{\boldsymbol{X}}), \tag{32}$$

with the weighting matrix $\boldsymbol{Q}_i$. In the sequel, the index i of the weighting matrix is dropped for the ease of notation. Evaluating (32) using (30) then yields

$$\begin{aligned} &\|\boldsymbol{Q}^{1/2}(\boldsymbol{W}_n - \hat{\boldsymbol{W}}_n)\|_F^2 = \\ &= \boldsymbol{b}^T(\boldsymbol{W}_n - \boldsymbol{1}\boldsymbol{m}^T - \boldsymbol{W}_o\boldsymbol{C})^T\boldsymbol{Q}(\boldsymbol{W}_n - \boldsymbol{1}\boldsymbol{m}^T - \boldsymbol{W}_o\boldsymbol{C})\boldsymbol{b}. \end{aligned} \tag{33}$$

Next, both $\boldsymbol{W}_n$ and $\boldsymbol{W}_o$ are referenced to their weighted centroids:

$$\boldsymbol{W}_n = \tilde{\boldsymbol{W}}_n + \boldsymbol{1}\boldsymbol{\mu}_n^T \tag{34}$$

with

$$\boldsymbol{\mu}_n^T = \frac{1}{s_q}\boldsymbol{q}^T\boldsymbol{W}_n. \tag{35}$$

and

$$\boldsymbol{W}_o = \tilde{\boldsymbol{W}}_o + \boldsymbol{1}\bar{\boldsymbol{w}}_o^T \tag{36}$$

with

$$\bar{\boldsymbol{w}}_o^T = \frac{1}{s_q}\boldsymbol{q}^T\boldsymbol{W}_o. \tag{37}$$

Thereby $\boldsymbol{q}$ is defined as the main diagonal of $\boldsymbol{Q}$ and s_q is the sum of all weights:

$$\boldsymbol{q} = \text{diag}(\boldsymbol{Q}), \quad s_q = \boldsymbol{1}^T\boldsymbol{q}.$$

This makes (33)

$$\begin{aligned} &\|\boldsymbol{Q}^{1/2}(\boldsymbol{W}_n - \hat{\boldsymbol{W}}_n)\|_F^2 = \boldsymbol{b}^T\{\tilde{\boldsymbol{W}}_n^T\boldsymbol{Q}\tilde{\boldsymbol{W}}_n + s_q(\boldsymbol{\mu}_n - \boldsymbol{m})(\boldsymbol{\mu}_n - \boldsymbol{m})^T - \\ &-2\boldsymbol{C}^T(\tilde{\boldsymbol{W}}_o^T\boldsymbol{Q}\tilde{\boldsymbol{W}}_n - s_q\bar{\boldsymbol{w}}_o\boldsymbol{m}^T + s_q\bar{\boldsymbol{w}}_o\boldsymbol{\mu}_n^T) + \boldsymbol{C}^T(\tilde{\boldsymbol{W}}_o^T\boldsymbol{Q}\tilde{\boldsymbol{W}}_o + s_q\bar{\boldsymbol{w}}_o\bar{\boldsymbol{w}}_o^T)\boldsymbol{C}\}\boldsymbol{b} = \min \end{aligned} \tag{38}$$

Collecting terms ($\boldsymbol{q}^T\tilde{\boldsymbol{W}}_o = \boldsymbol{0}$ and $\boldsymbol{q}^T\tilde{\boldsymbol{W}}_n = \boldsymbol{0}$) and using the abbreviation $\boldsymbol{c}_o = \boldsymbol{C}^T\bar{\boldsymbol{w}}_o$ leads to

$$\begin{aligned} &\|\boldsymbol{Q}^{1/2}(\boldsymbol{W}_n - \hat{\boldsymbol{W}}_n)\|_F^2 = \boldsymbol{b}^T\{(\tilde{\boldsymbol{W}}_n - \tilde{\boldsymbol{W}}_o\boldsymbol{C})^T\boldsymbol{Q}(\tilde{\boldsymbol{W}}_n - \tilde{\boldsymbol{W}}_o\boldsymbol{C}) + \\ &+s_q[(\boldsymbol{\mu}_n - \boldsymbol{m} - \boldsymbol{c}_o)(\boldsymbol{\mu}_n - \boldsymbol{m} - \boldsymbol{c}_o)^T]\}\boldsymbol{b} = \min \end{aligned} \tag{39}$$

The minimisation of (39) now has to be carried out with respect to $\boldsymbol{C}$, $\boldsymbol{m}$ and $\boldsymbol{b}$. The centroid $\boldsymbol{m}$ only appears in the second term of (39) which is a positive semidefinite expression. Therefore, $\boldsymbol{m}$ has to be chosen such that the second term in (39) vanishes:

$$\boldsymbol{m} = \boldsymbol{\mu}_n - \boldsymbol{C}^T\bar{\boldsymbol{w}}_o. \tag{40}$$

The first term in (39) is independent of $\boldsymbol{m}$ and has to be minimised separately. Its minimisation can be re-written as

$$\begin{aligned} &\boldsymbol{b}^T(\tilde{\boldsymbol{W}}_n - \tilde{\boldsymbol{W}}_o\boldsymbol{C})^T\boldsymbol{Q}(\tilde{\boldsymbol{W}}_n - \tilde{\boldsymbol{W}}_o\boldsymbol{C})\boldsymbol{b} = \\ &\|\boldsymbol{Q}^{1/2}(\tilde{\boldsymbol{W}}_n - \tilde{\boldsymbol{W}}_o\boldsymbol{C})\boldsymbol{b}\|_F^2 = \min_{\boldsymbol{C},\boldsymbol{b}}. \end{aligned}$$

The above Frobenius norm can be expanded in the following way:

$$\|\boldsymbol{Q}^{1/2}(\tilde{\boldsymbol{W}}_n - \tilde{\boldsymbol{W}}_o\boldsymbol{C})\boldsymbol{b}\|_F^2 =$$

$$\|\boldsymbol{Q}^{1/2}(\tilde{\boldsymbol{w}}_{n,1} - \tilde{\boldsymbol{W}}_o \boldsymbol{c}_1)b_1\|_2^2 + \|\boldsymbol{Q}^{1/2}(\tilde{\boldsymbol{w}}_{n,2} - \tilde{\boldsymbol{W}}_o \boldsymbol{c}_2)b_2\|_2^2 + \dots$$

where $\tilde{\boldsymbol{w}}_{n,j}$ and $\boldsymbol{c}_j$ denote the j-th column vectors of $\tilde{\boldsymbol{W}}_n$ and $\boldsymbol{C}$, respectively and b_j denotes the j-th element of the vector $\boldsymbol{b}$.
The minimisation of the Frobenius norm with respect to the two arguments $\boldsymbol{C}, \boldsymbol{b}$ can now be split into two subtasks:

1. The norms $\|\boldsymbol{Q}^{1/2}(\tilde{\boldsymbol{w}}_{n,j} - \tilde{\boldsymbol{W}}_o \boldsymbol{c}_j)\|_2^2$ of every single column vector have to be minimised which determines every column vector $\boldsymbol{c}_j$ of $\boldsymbol{C}$ independently. This can be summarised in one matrix operation:

$$Tr\{(\tilde{\boldsymbol{W}}_n - \tilde{\boldsymbol{W}}_o \boldsymbol{C})^T \boldsymbol{Q}(\tilde{\boldsymbol{W}}_n - \tilde{\boldsymbol{W}}_o \boldsymbol{C})\} = \min_{\boldsymbol{C}}.$$

$$\frac{\partial Tr}{\partial \boldsymbol{C}} = -2\tilde{\boldsymbol{W}}_o^T \boldsymbol{Q}(\tilde{\boldsymbol{W}}_n - \tilde{\boldsymbol{W}}_o \boldsymbol{C}) = 0 \tag{41}$$

$$\boldsymbol{C} = (\tilde{\boldsymbol{W}}_o^T \boldsymbol{Q} \tilde{\boldsymbol{W}}_o)^{-1} \tilde{\boldsymbol{W}}_o^T \boldsymbol{Q} \tilde{\boldsymbol{W}}_n \tag{42}$$

   Note that (42) is the solution of the overdetermined system $\tilde{\boldsymbol{W}}_o \boldsymbol{C} = \tilde{\boldsymbol{W}}_n$ where every row is weighted individually by the elements of $\boldsymbol{q}$.

2. Vector $\boldsymbol{b}$ becomes the unit eigenvector corresponding to the minimal eigenvalue of $(\tilde{\boldsymbol{W}}_n - \tilde{\boldsymbol{W}}_o \boldsymbol{C})^T \boldsymbol{Q}(\tilde{\boldsymbol{W}}_n - \tilde{\boldsymbol{W}}_o \boldsymbol{C})$.

Remark: The special case when only $y(k)$ is noisy leads to a partitioning $\boldsymbol{W}_n = [\boldsymbol{y}]$ and $\boldsymbol{W}_o = [\boldsymbol{X}]$ which essentially makes (42) a conventional LS solution. If, on the other hand all components are noisy then $\boldsymbol{C}$ vanishes and one obtains the TLS solution.

3.3 The GTLS residual

In this section a statistical criterion is derived that enables the statistical assessment of the residual error. This is an important prerequisite in a local model network with data-based partitioning. The GTLS residual is used to discriminate between unsystematic errors from measurement noise and systematic errors from truncation.
If the weights contained in $\boldsymbol{Q}_i$ are referred to a local model η_i a *class-conditional* GTLS-residual can be defined:

$$r(\boldsymbol{w}^T(k), \eta_i) = [\tilde{\boldsymbol{w}}_n^T(k) - \tilde{\boldsymbol{w}}_o^T(k)\boldsymbol{C}]\boldsymbol{b} \tag{43}$$

The argument η_i for r was instanced in order to emphasize that for the given model (indexed by i) all parameters ($\boldsymbol{b}$, $\boldsymbol{m}$ and $\boldsymbol{C}$) and consequently also the residual r essentially depend on the weight $q_i(k)$ at which every single training data record influences the parameters. Given the weights $q_i(k)$ the weighted GTLS parameter estimation (32) can be formally written as

$$\sum_{k=1}^{N} q_i(k)\, r(\boldsymbol{w}^T(i), \eta_i) \frac{\partial r}{\partial \boldsymbol{\theta}_i} = 0. \tag{44}$$

Note that for the special case that only $y(k)$ is noisy (43) simply reduces to the prediction error and (44) results in WLS parameter estimation.
If the noise signals in $\boldsymbol{W}_n$ are Gaussian with variance one and zero cross-correlations then the GTLS-residual r also follows a Gaussian distribution with zero mean and unit variance. Consequently, a *class-conditional residual distribution function* can be defined:

$$p(\boldsymbol{w}^T(k), \eta_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{r^2(\boldsymbol{w}^T(k), \eta_i)}{2}\right), \tag{45}$$

with r from (43).
The residual distribution function always refers to a linear model with parameters obtained from weighted GTLS with the weights chosen according to the class η_i.
It describes the probability density of a data record $(y(k), \boldsymbol{x}(k))$, belonging to model η_i and having a GTLS residual r. For an increasing number of data records used for training the empirical distribution of the residual should follow (45) if the model is actually linear or - more generally - if the data can be described by the chosen model structure. If there are nonlinearities that cannot be described by the model then the actual residual distribution will deviate from (45) in size and shape.
For illustration Fig. 8 compares the histograms of actual GTLS residual distributions to the ideal distribution (45). The upper figure shows an adverse situation which is caused by nonlinearities which cannot be properly modeled by the GTLS model structure (note the outlier at $r = 8$) whereas the lower figure shows a result, where the training data originate from a suitable (linear) process.

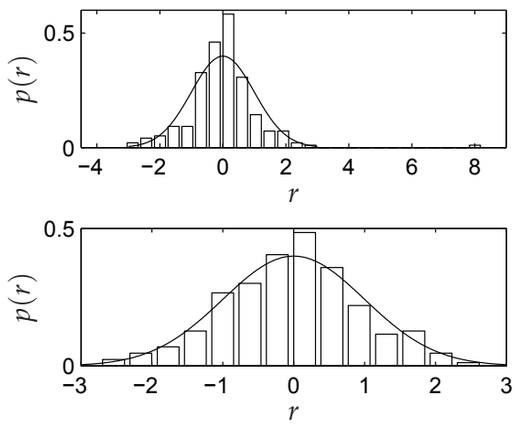


Fig. 8. Residual distribution for an adverse (top) and proper (bottom) model structure

4. Illustrative example

A simulation example is chosen to demonstrate the applicability and the benefits of the proposed concepts. For that purpose a nonlinear dynamic MISO system with three inputs was chosen from Treichl et al. (2002), where it is used to validate another MISO identification scheme. The block diagram is depicted in Fig. 9.
The transfer functions $F_1(z^{-1})$, $F_2(z^{-1})$ and $F_3(z^{-1})$ and the nonlinear static characteristics are chosen very similar to Treichl et al. (2002):

$$\begin{aligned} F_1(z^{-1}) &= \frac{0.1092z^{-1}+0.09552z^{-2}}{1-1.605z^{-1}+0.6703z^{-2}} \\ F_2(z^{-1}) &= \frac{0.2183z^{-1}+0.191z^{-2}}{1-1.605z^{-1}+0.6703z^{-2}} \\ F_3(z^{-1}) &= \frac{0.1092z^{-1}+0.09552z^{-2}}{1-1.605z^{-1}+0.6703z^{-2}} \end{aligned}$$

$$\begin{aligned} \mathcal{NL}_1: \quad v_1 &= -\tfrac{1}{2}u_1 + \tfrac{1}{2}u_1^2 - u_1^3 \\ \mathcal{NL}_2: \quad v_2 &= -\tfrac{1}{2}u_2 - u_2^2 + \tfrac{1}{2}u_3^3 \\ \mathcal{NL}_3: \quad v_3 &= \tfrac{1}{2} - \tfrac{1}{2}u_3 - u_2^3 \end{aligned}$$

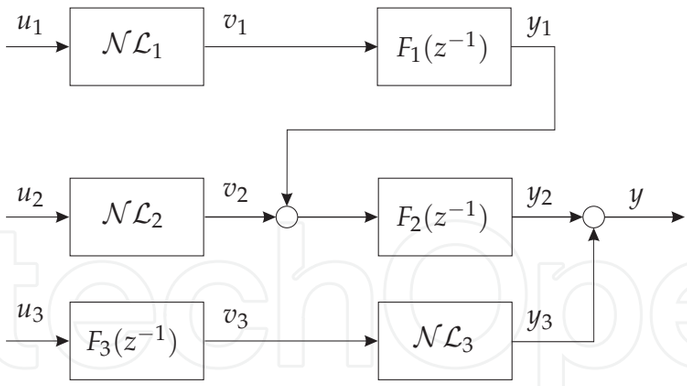


Fig. 9. Structure of the nonlinear MISO model

	LS		GTLS	
Data	R^2	MSE	R^2	MSE
identification	0.97924	0.13289	0.99193	0.051672
validation	0.97706	0.15699	0.99331	0.045785

Table 1. Simulation results

In this example, the input u_1 and the output are corrupted by Gaussian noise. For the excitation of the system, i. e. for the simulation of the training and generalisation data record, respectively, APRB-signals are selected. Their amplitudes and bandwidths are designed such that the whole static and dynamic operating range of the MISO system is covered.
In table 1 a comparison of the simulation results with LS and GTLS parameter estimation with eight local models is presented. It is clearly visible that the performance model with GTLS parameter estimation and partitioning based on the GTLS residual is considerably better than the LS model. This result is reflected in Fig. 10 where the autocorrelation function of the prediction error of the validation data record for LS and GTLS parameter estimates is depicted. The validation data record is chosen to be noise-free in this example to separate the model error from measurement error.
In Fig. 11 the simulation results with validation data are presented.

5. Conclusion

In this chapter the problem of noise in measured data in nonlinear system identification is addressed. First, a robust and efficient partitioning strategy using an EM algorithm is proposed. Second, weighted TLS and GTLS algorithms are presented which yield consistent parameter estimates of the local model parameters when some (GTLS) or all (TLS) input channels are turn out as special case of GTLS.
The GTLS residual is defined, which allows the statistical assessment of the residual error. This is an important prerequisite in a local model network with data-based partitioning.
The benefits of the proposed concepts are demonstrated by means of a simulation example. The performance of the resulting nonlinear model with local parameters estimated by

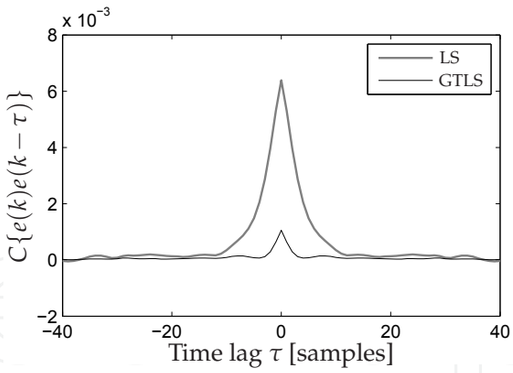


Fig. 10. Autocorrelation function of the prediction error for validation data with LS and GTLS parameter estimates, respectively.

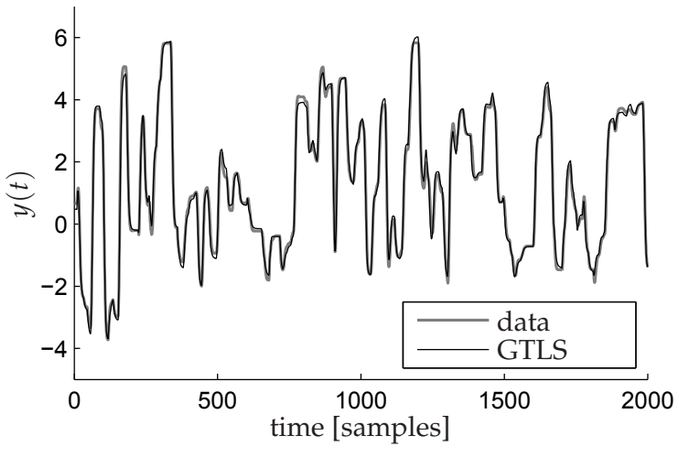


Fig. 11. Comparison of validation data to nonlinear models with GTLS parameter estimates

weighted GTLS is a product both of the parameter estimation itself and the associated residual used for the partitioning process.

6. References

Abonyi, J., Babuska, R. & Szeifert, F. (2002). Modified Gath-Geva Fuzzy Clustering for Identification of Takagi-Sugeno Fuzzy Models, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 32, IEEE, pp. 612–621.

Breiman, L., Friedman, J., Olshen, R. & Stone, C. (1984). *Classification and Regression Trees*, Wadsworth and Brooks, Monterey, CA.

De Groen, P. (1996). An introduction to total least squares, *Nieuw Archief Voor Wiskunde* **14**(2): 237–254.
URL: *http://www.citebase.org/cgi-bin/citations?id=oai:arXiv.org:math/9805076*

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum Likelihood From Incomplete Data Via The EM Algorithm, *Journal of the Royal Statistical Society, Series B* **39**(1): 1–38.

Golub, G. & Loan, C. V. (1980). An analysis of the Total Least Squares problem, *SIAM J. Numer. Anal.* **17**(3): 883–893.

Hametner, C. & Jakubek, S. (2007). Neuro-fuzzy modelling using a logistic discriminant tree, *American Control Conference, 2007. ACC '07* pp. 864–869.

Han, S., Kim, J. & Sung, K. (1996). Extended Generalized Total Least Squares Method for the Identification of Bilinear Systems, *IEEE Transactions on Signal Processing*, Vol. 44, IEEE, pp. 1015 – 1018.

Heij C., S. W. (1999). Consistency Of System Identification By Global Total Least Squares, *Automatica* **35**: 993–1008.

Huffel, S. V. & Zha, H. (1991). The Restricted Total Least Squares Problem: Formulation, Algorithm, and Properties, *SIAM Journal on Matrix Analysis and Applications* **12**(2): 292–309.
URL: *http://link.aip.org/link/?SML/12/292/1*

Jakubek, S. & Keuth, N. (2006). A Local Neuro-Fuzzy Network for High-Dimensional Models and Optimisation, *Engineering Applications of Artificial Intelligence* **19**: 705–717.

Johansen, T. A., Shorten, R. & Murray-Smith, R. (2000). On the Interpretation and Identification of Dynamic Takagi-Sugeno Fuzzy Models, *IEEE Transactions on Fuzzy Systems* **8**(3): 297–313.

Markovsky, I. & Huffel, S. V. (2007). Overview of total least-squares methods, *Signal Process.* **87**(10): 2283–2302.

Markovsky, I., Willems, J., Van Huffel, S., De Moor, B. & Pintelon, R. (2005). Application of Structured Total Least Squares for System Identification and Model Reduction, *IEEE Transactions On Automatic Control* **50**: 1490–1500.

Nayak, A., Trucco, E. & Thacker, N. A. (2006). When are Simple LS Estimators Enough? An Empirical Study of LS, TLS, and GTLS, *Int. J. Comput. Vision* **68**(2): 203–216.

Nelles, O. (2002). *Nonlinear System Identification*, 1st edn, Springer Verlag.

Paige, C. C. & Wei, M. (1993). Analysis of the generalized Total Least Squares Problem $AX = B$ when some columns of A are free of error, *Numerical Mathematics* **65**: 177–202.

Roorda, B. (1995). Algorithms for Global Total Least Squares Modelling of Finite Multivariable Time Series, *Automatica* **31**(3): 391 – 404.

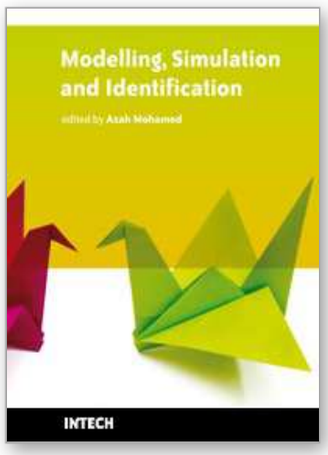
Soderstrom, T. (2007). Errors-in-variables methods in system identification, *Automatica* **43**(6): 939–958.
URL: *http://www.sciencedirect.com/science/article/B6V21-4NH6NBT-2/2/74cb4630fabb1a3fbf568dbe52a8f8b6*

Theodoridis, S. & Koutroumbas, K. (1999). *Pattern Recognition*, Academic Press, San Diego, CA, USA.

Treichl, T., Hofmann, S. & Schroder, D. (2002). Identification of nonlinear dynamic MISO systems with orthonormal base function models, *Proceedings of the 2002 IEEE International Symposium on Industrial Electronics*, Vol. 1, pp. 337– 342.

Van Huffel, S. & J., V. (1989). Analysis and Propoerties of the generalized Total Least Squares Problem $AX = B$ when some or all columns of A are subject to Error, *SIAM Journal on Matrix Analysis and Applications* **10**(3): 294 – 315.

Vandersteen, G. (1998). On the Use of Compensated Total Least Squares in System Identification, *IEEE Transactions on Automatic Control*, Vol. 43, IEEE, pp. 1436 – 1441.



Modelling, Simulation and Identification

Edited by Azah Mohamed

ISBN 978-953-307-136-7
Hard cover, 354 pages
Publisher Sciyo
Published online 18, August, 2010
Published in print edition August, 2010

Modeling, simulation and identification has been actively researched in solving practical engineering problems. This book presents the wide applications of modeling, simulation and identification in the fields of electrical engineering, mechanical engineering, civil engineering, computer science and information technology. The book consists of 17 chapters arranged in an order reflecting multidimensionality of applications related to power system, wireless communication, image and video processing, control systems, robotics, soil mechanics, road engineering, mechanical structures and workforce capacity planning. New techniques in signal processing, adaptive control, non-linear system identification, multi-agent simulation, eigenvalue analysis, risk assessment, modeling of dynamic systems, finite difference time domain modeling and visual feedback are also presented. We hope that readers will find the book useful and inspiring by examining the recent developments in the applications of modeling, simulation and identification.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Christoph Hametner and Stefan Jakubek (2010). Nonlinear System Identification through Local Model Approaches: Partitioning Strategies and Parameter Estimation, Modelling, Simulation and Identification, Azah Mohamed (Ed.), ISBN: 978-953-307-136-7, InTech, Available from: http://www.intechopen.com/books/modelling--simulation-and-identification/nonlinear-system-identification-through-local-model-approaches-partitioning-strategies-and-parameter

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen