Discussion on “On the Use of Minimal Parametrizations in Multivariable ARMAX Identification” by R. P. Guidorzi

Tomas McKelvey

Department of Electrical Engineering
Linköping University, S-581 83 Linköping, Sweden
WWW: http://www.control.isy.liu.se
Email: tomas@isy.liu.se

March 1999

Report no.: LiTH-ISY-R-2132

Technical reports from the Automatic Control group in Linköping are available by anonymous ftp at the address ftp.control.isy.liu.se. This report is contained in the pdf-file 2132.pdf.
Discussion on “On the Use of Minimal Parametrizations in Multivariable ARMAX Identification” by R. P. Guidorzi

Tomas McKelvey
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping, Sweden
email: tomas@isy.liu.se

May 1998

This discussion section will focus on three issues:

1. **Minimal parametrizations** Is a minimal parametrization a necessity for a successful identification result?

2. **The structural issue** Does the choice of structural indices influence the identification result?

3. **Gauss-Newton optimization** How should the GN-search direction be calculated in a numerically stable way?

**Minimal Parametrizations**

System identification fundamentally deals with selecting a linear system based on some metric which depends on the data. In the paper the prediction error method is employed which corresponds to minimizing the sum of the square of the prediction errors. In order to execute this selection a model structure and parametrization needs to be defined which forms a mapping $M$ from the space of real parameters $\mathcal{R}^l$ to the space of prediction models $\Sigma$. If this mapping $M$ is injective the parametrization is known as minimal, i.e. a prediction model corresponds to at most one point in the parameter space. A fundamental result states that for multi output ARMAX systems there exists no injective mapping $M$ which covers all linear models of a given McMillan degree. Hence a set of parametrizations is needed, $M_i$, $i = 1, \ldots$. In the paper one such set is presented which is based on polynomial input output models. In its original form it is not suitable for predictions but a clever reformulation, see section 2.1, enables a minimal ARMAX parametrization of a predictor.

Is it necessary to use a minimal parametrization? Consider the state-space representations, see equation (1), and let all matrix elements in $A, B, C, K$ represent the parametrization of the linear model [4]. Trivially this parametrization is a surjective map $M$ between the parameter space $\mathcal{R}^l$ and the space of linear
models $\Sigma$, i.e. every linear system of McMillan degree $n$ has infinitely many representations in the parameter space. We call this an overparametrized model structure. It is important to notice that the overparametrized model structure and the set of minimal ARMAX parametrizations describes the same set of linear systems (1). Hence the selection of models from this common set by some metric does not depend on the parametrization used, minimal or not. The identification result will, if we disregard numerical issues, be identical. Also notice that from a stochastic point of view the statistical properties of the identified model (such as the variance of the transfer function) is also invariant with respect to the parametrization used [7]. The advantage of an overparametrization over a minimal one is that it is sufficient to consider one parametrization as well as the possibility to use a state-space basis which is well conditioned, e.g. balanced forms. An apparent negative property of the overparametrized models is the increased parameter dimension. However this can be reduced by a tridiagonal parametrization [5] or dynamically, during the optimization, form a minimal parametrization [6].

The Structural Issue

Each minimal ARMAX parametrization $M_i$ corresponding to a certain structural index $\nu$ is generic, i.e. the set of systems which cannot be described by the parametrization is of measure zero. Hence, given an arbitrary system, any of the ARMAX parametrizations $M_i$ will, w.p. 1, exactly be able to model it, [3]. This fact can lead you to the conclusion, which the author does, that the structural issue is only a theoretical one without practical identification implications. To show the converse first the results of a multivariable identification problem is presented followed by an analysis of the underlying reason which leads us to the third and last issue.

Identification of a pendulum system

The problem is to identify an output-error model of an educational pendulum system from measured data using PEM. The system consists of a pendulum attached to a moving cart driven by a DC-motor. The input is the Voltage to the DC-motor and the 4 measurements are, cart position and velocity, pendulum angle and velocity. In figure 1 the criterion function $V(\theta)$ is plotted as a function of the number of Gauss-Newton iterations for two different parametrizations, the observable state-space structure [3, p. 119] which is a minimal parametrization and a full state-space parametrization [6]. The same Gauss-Newton optimization algorithm has been used for both parametrizations. Clearly there is a fundamental difference in the speed of convergence which is due to the type of parametrization used.

Gauss-Newton optimization

The quadratic criterion function to be minimized, see equations (26-31) of the paper, can be written as 

$$V(\theta) = e^o(\theta)^T e^o(\theta)$$

where $e^o(\theta)$ is a $Nm-$dimensional vector of all prediction errors. Denote by $-H_\phi(\theta)$ the Jacobian matrix of dimension $Nm \times l$ associated with $e^o(\theta)$. The
Gauss-Newton search direction $\gamma(\theta)$ can be formulated as the solution to the following Least-Squares (LS) problem

$$\gamma(\theta) = \arg\min_x \| e^\circ(\theta) - H_{\psi}(\theta)x \|^2$$

and the parameter update is

$$\theta^{k+1} = \theta^k + \delta_k \gamma(\theta^k)$$

where $\delta_k$ is adjusted to ensure convergence (damped Gauss-Newton). The Gauss-Newton direction can be determined by straightforward solution of the quadratic problem via the normal equations, c.f. equation (33) in the paper

$$\gamma(\theta) = \left[ H_{\psi}(\theta)^T H_{\psi}(\theta) \right]^{-1} H_{\psi}^T(\theta)e^\circ(\theta)$$

Since the Gauss-Newton step is the solution to a Least-Squares problem its relative accuracy is proportional to the square of the condition number of $H_{\psi}(\theta)$ [2]. In figure 2 the development of the condition number is depicted for the pendulum identification example. Clearly the minimal parametrization has a higher condition number. The relative sensitivity of the GN-direction will be higher and the search direction becomes less accurate. This explains the slow convergence.

Slow convergence is thus correlated with an ill-conditioned $H_{\psi}(\theta)$ matrix. One possibility to improve conditioning is to use a pseudo-inverse [2] of $H_{\psi}$ when solving for the GN-direction or to add a scaled identity matrix to $H_{\psi}$ (Levenberg-Marquardt) [1]. When the pseudo-inverse is employed the singular
values of $H_\psi$ matrix is determined an all less than $\zeta \bar{\sigma}$ are set to zero, where $\bar{\sigma}$ is the largest singular value of $H_\psi$ and $\zeta$ is a tunable constant. When pseudo-inverse technique is employed with the minimal parametrization and a choice of $\zeta = 10^{-4}$, we get the results shown in figures 3 and 4. An improved convergence to a point near the model obtained with the full parametrization can be noticed. The condition number of (the unmodified) $H_\psi$ has also greatly improved which indicates that the parameter path towards convergence has been shifted towards a path where $H_\psi$ is better conditioned.

**Summary**

When performing practical identification the structural issue often cannot be neglected since the conditioning of the parameterization influence the convergence of the optimization. Solutions are to carefully monitor the conditioning of the associated LS problem or to use alternative parametrizations (minimal or non-minimal) which give better conditioning.

**References**

Figure 3: Identification results during Gauss-Newton search using modified and non-modified GN-direction. The graph shows the evolution of the criterion function $V(\theta)$ during the minimization.

Figure 4: Identification results during Gauss-Newton search using modified and non-modified GN-direction. The graph illustrates the development of the condition number of $H_\psi$, the Jacobian matrix during the optimization.


