The Role of Parametrizations in Identification of Linear Systems

Manfred Deistler
Institute for Econometrics, Operations Research and System Theory
Vienna University of Technology
Argentinierstr. 8, A-1040 Wien
Tel.: ++43 +1 58801 11911, Fax.: ++43 +1 58801 11999
email: Manfred.Deistler@tuwien.ac.at

1 Introduction

In identification the problem is to attach to every string of data of the form $y_1, \ldots, y_T$; $y_t \in \mathbb{R}^s$, a system from an a priori specified model class. Ususally the model class is described by a space of free parameters. In the fully automatized case, the system (or its free parameters) is attached to the data by a function, ψ say. If the data are assumed to be generated by an underlying stochastic process $(y_t | t \in \mathbb{Z})$ (called the data generating process, DGP) and if ψ is measurable, then ψ is an estimator and the identification problem is an estimation problem. The special features of system identification arise from the rather complicated relation between external behavior, internal system parameters and free parameters for a given model class.

For simplicity here we consider linear, finite dimensional, time-invariant, causal and stable systems only, where in addition the only inputs are unobserved white noise (ε_t). We discuss state space and ARMA forms. The state space systems are assumed to be in innovations representation:

$$\begin{aligned} x_{t+1} &= Ax_t + B\varepsilon_t \\ y_t &= Cx_t + \varepsilon_t \end{aligned}$$

Here y_t are the s-dimensional outputs, x_t the n-dimensional states and $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times s}$ and $C \in \mathbb{R}^{s\times n}$ are the matrices of system parameters. The noise parameters are $\Sigma = \mathbb{E}\varepsilon_t\varepsilon_t'$. We impose the stability condition

$$|\lambda_{max}(A)| < 1$$

and the miniphase condition

$$|\lambda_{max}(A - BC)| \leq 1$$

where λ_{max} denotes an eigenvalue of maximal modulus. In addition we assume

$$\Sigma > 0.$$

The transfer function is of the form

$$k(z) = \sum_{j=1}^{\infty} K_j z^j + I, \quad K_j = CA^{j-1}B, \quad z \in \mathbb{C}$$

For ARMA systems we use the notation

$$a(z)y_t = b(z)\varepsilon_t$$

where z here is the backwardshift,

$$a(z) = \sum_{j=0}^{p} A_j z^j, \quad b(z) = \sum_{j=0}^{q} B_j z^j$$

where $A_j, B_j \in \mathbb{R}^{s\times s}$, and we assume the stability condition

$$\det\, a(z) \neq 0 \quad |z| \leq 1$$

and the miniphase condition

$$\det\, b(z) \neq 0 \quad |z| < 1$$

In addition we assume $A_0 = B_0$. Then the transfer function is given by

$$k(z) = \sum_{j=1}^{\infty} K_j z^j + I = a^{-1}(z)b(z)$$

We may distinguish parametric, nonparametric and semi-nonparametric estimation. In parametric estimation the parameter spaces are finite dimensional (and the parameters determine the probability law of the observations, however we will disregard this second point here). In non-parametric estimation e.g. functions are estimated, such as transfer functions without degree restrictions or even rationality assumption. In semi-nonparametric estimation, the original parameter space, Θ say, may not be finite dimensional, but Θ is broken into finite dimensional bits Θ_α say. Then estimation consists of two parts, namely estimation of α (model selection) and estimation of the finite dimensional parameters, θ say, in Θ_α. This is the approach considered here. In this context the identification problem may be decomposed into the following three modules:

- Structure Theory: Here an idealized indentification problem is treated, as we commence from the external behaviour, as described by the transfer function in the linear case (and from Σ) rather than from data, to obtain the internal parameters.
- Estimation of real valued parameters for a selected model (sub -) class Θ_α, i.e. estimation of $\theta \in \Theta_\alpha$.
- Model selection, i.e. estimation of α. Usually α is a multi-index of integers.

Let U_A denote the set of all rational $s \times s$ transfer functions, having no poles for $|z| \leq 1$ and no zeros for $|z| < 1$ and satisfying $k(0) = I$ and let $U \subset U_A$ denote the a priori specified class of transfer functions. Unless the contrary is stated explicitely, we take $U = U_A$. Now the way how U is broken into bits, U_α say, and how the these bits are parameterized is of significant influence for the properties of identification procedures as will be demonstrated in this contribution.

2 Parametrizing Classes of Linear Systems

The description of a transfer function $k \in U$, in general, is performed in three steps:

- In the first step U is broken into bits U_α, $\alpha \in I$, say and integer valued parameters α are attached to k, such that $k \in U_\alpha$.
- In the second step, for a given U_α, k is described by a finite dimensional vector of (real valued) system parameters $\tau \in T_\alpha$, say.
- Since, in general, for given T_α, not all system parameters are free, we may describe a system by its free parameters $\tilde{\tau} \in \tilde{T}_\alpha$ say.

Let T_A denote the set of all state space systems (A, B, C), where s is fixed but n is variable, satisfying our assumptions.

Let

$$\begin{aligned} \pi : T_A &\rightarrow U_A \\ \pi(A, B, C) &= C(zI - A)^{-1}B + I \end{aligned}$$

denote the mapping attaching transfer functions to system parameters. For given U_α we only consider T_α such that $\pi(T_\alpha) = U_\alpha$ and T_α can be embedded in a finite dimensional Euclidean space, $\mathbb{R}^{d_\alpha}$ say. In this case U_α is called finite dimensional. Clearly, if π restricted to T_α is injective, then we have identifiability, and thus there exists a *parametrizing mapping*

$$\psi_\alpha : U_\alpha \rightarrow T_\alpha.$$

We use the same notation for ARMA systems. Further, let

$$\rho : T_\alpha \rightarrow \tilde{T}_\alpha \in \mathbb{R}^{\tilde{d}_\alpha}$$

denote the mapping attaching free parameters to system parameters; we assume that ρ is bijective and that $\tilde{T}_\alpha$ contains a nonvoid set open in $\mathbb{R}^{\tilde{d}_\alpha}$.

By $M(n) \subset U_A$ we denote the set of all transfer functions of order n, by $S(n) \subset T_A$ we denote the set of all state space systems (A, B, C) with state dimension n and by $S_m(n) \subset S(n)$ the subset of all minimal (A, B, C). Finally, U_A is endowed with the so called pointwise topology which corresponds to the relative topology in the product space $(\mathbb{R}^{s \times s})^{\mathbb{N}}$ for the $(K_j | j \in \mathbb{N})$.

The following approaches for parametrizing classes of linear systems are considered in the lecture.

1. Echelon and balanced canonical forms as the most important examples for canonical forms (see e.g. [4] and [1]).
2. The description of the manifold $M(n)$ by the local coordinates described by dynamical indices (see e.g. [4]).
3. The ARMA parametrization with prescribed column degrees (see e.g. [2]).

4. The ARMA parametrization with prescribed polynomial zeros (see e.g. [4]).

5. The ARMA parametrization commencing from writing k as $c^{-1}p$, where c is the least common denominator polynomial for k and where the degrees of c and p serve as integer valued parameters.

6. The description of $M(n)$ and $\bar{M}(n)$ by $S_m(n)$ and $S(n)$, respectively, which is non-identifiable.

3 Estimating Integer Valued and Real Valued Parameters

As a prototype procedure the following is considered:

For given U_α (Gaussian) maximum likelihood estimation is performed: Let $L_T(k, \Sigma)$ denote the (negative log-) likelihood as a function of the transfer function k and the innovation variance Σ; then the maximum likelihood estimators (MLE's) are obtained as

$$\hat{k}_T, \hat{\Sigma}_T = \arg\min_{U_\alpha \times \underline{\Sigma}} L_T(k, \Sigma)$$

where $\underline{\Sigma} = \{\Sigma | \Sigma > 0\}$. Under general conditions (see [4]), the following coordinate free consistency result holds: $\hat{k}_T \to k_0$ a.s., $\hat{\Sigma}_T \to \Sigma_0$ a.s., where $k_0 \in \bar{U}_\alpha$ and Σ_0 denote the true values, corresponding to the DGP. The corresponding parameter estimators then may be defined as $\hat{\tau}_T = \psi_\alpha(\hat{k}_T)$ (if $\hat{k}_T \in U_\alpha$; note that $\hat{k}_T$ may be in $\bar{U}_\alpha - U_\alpha$) or $\tilde{\tau} = \rho(\hat{\tau}_T)$.

In a second step model selection is performed by minimizing a criterion of the form

$$A(\alpha) = \log \det \hat{\Sigma}_T(\alpha) + \tilde{d}_\alpha \cdot c(T) \cdot T^{-1}, \quad \alpha \in I_1 \quad (1)$$

where $I_1 \subset I$, $\hat{\Sigma}_T(\alpha)$ is the MLE of Σ_0 over $U_\alpha \times \underline{\Sigma}$, $\tilde{d}_\alpha$ is the dimension of $\tilde{T}_\alpha$ and $c(T)$ is a prescribed function; choosing $c(T)$ as $c{\cdot}\log(T)$, $c \geq 1$ gives the BIC criterion.

As a general remark note that in comparing different decompositions into bits there is a tradeoff between the quality in estimating integer valued parameters versus estimating real valued parameters.

It should also be noted that the model selection step adds uncertainty to the estimators of the real valued parameters in a highly non-trivial way (see [5]).

Now for consistency of parameter estimation the following desirable properies of parametrizations may be required:

1. T_α is identifiable

2. The parametrizing mapping $\psi_\alpha : U_\alpha \to T_\alpha$ is a homeomorphism.

3. The mapping ρ is a homeomorphism.

4. U_α is open relative to its closure $\bar{U}_\alpha$.

If $\hat{k}_T \to k_0$, $k_0 \in U_\alpha$ holds, then the openness of U_α in $\bar{U}_\alpha$ implies that $\psi_\alpha(\hat{k}_T) = \hat{\tau}_T$ exists from a certain T_0 onwards and then, by the continuity requirements (2.) and (3.) we have consistency for the parameters, i.e. $\hat{\tau}_T \to \tau_0 = \psi_\alpha(k_0)$ a.s. and $\rho(\hat{\tau}_T) = \tilde{\tau}_T \to \tilde{\tau}_0 = \rho(\tau_0)$ a.s.

Another desirable property is that

5. $\cup_{\alpha \in I} U_\alpha = U_A$

holds, i.e. that every system can be described by a suitable choice of α.

In the lecture the six approaches for parametrizing linear systems mentioned above are analyzed with respect to these desirable properties (compare [3]). In addition the respective boundary points and the behavior of the estimators at these boundary points are discussed. For estimation of the integers α by a criterion of the form (1), to be more specific for consistent estimation of α by BIC, the sets U_α, $\alpha \in I_1$ have to satisfy additional requirements. Such a requirement is

6. The sets U_α, $\alpha \in I_1$ are *closure nested*. This means that there exists an ordering on I_1 such that $\alpha \geq \beta$ implies $\bar{U}_\alpha \supset U_\beta$

The $M(n), n = 0, 1, \ldots$ are an example for such a closure nestedness. The parametrizations (1.) - (6.) will also be discussed with respect to this property in the lecture.

References

[1] Bauer, D. and Deistler, M. (1999). Balanced Canonical Forms for System Identification. IEEE Trans. Autom. Control, 44, 1118-1131.

[2] Deistler, M. (1983). The Properties of the Parametrization of ARMAX systems and their Relevance for Structural Estimation and Dynamic Specification. Econometrica, 51, no.4, 1187-1207.

[3] Deistler, M. and Wang, L., (1989). The Common Structure of Parametrizations for Linear Systems. Linear Algebra and its Applications 122/123/124, 921-941.

[4] Hannan, E. J. and Deistler, M., (1998). The Statistical Theory of Linear Systems. John Wiley & Sons, New York.

[5] Leeb, H. and Pötscher, B. M. (2000). The Finite Sample Distribution of Post-Model-Selection Estimation, and Uniform Versus Non-Uniform Approximations. Mimeo, Institute for Statistics and Decision Support Systems, University Vienna.