Least squares optimal identification of LTI SISO dynamical models is an eigenvalue problem

Abstract Workshop 'Vistas in Control', 50th birthday Automatic Control Library, ETH Zurich

Bart De Moor, www.bartdemoor.be
STADIUS, ESAT, KU Leuven, Belgium
+32 475 28 70 52 ; Bart.DeMoor@kuleuven.be

The last 50 years, we have witnessed tremendous progress in model-based control design and applications. Both the solution to the steady state LQR problem and the Kalman filter, ultimately reduce to an eigenvalue problem (which is hidden in the matrix Riccati equations involved). The same observation applies to the H-infty framework, with different Riccati equations.

But what about the models? Are they optimal in any sense, as often they are obtained from some non-linear optimization method? We will show how least squares optimal models for LTI SISO systems, derive from an eigenvalue problem, hence achieving an important landmark on optimal identification for this type of models. We focus on LTI SISO models, with at most one observed input and/or output data record, and at most one 'unobserved' (typically assumed to be white) noise input (called 'latency'). In the most general case, also the inputs and outputs are corrupted by additive measurement noise (called 'misfit'). This class of models covers a lot of special cases, like AR, MA, ARMA, ARMAX, OE (Output-Error), Box-Jenkins, EIV (Errors-in-Variables), dynamic total least squares, and also includes models that up to now have not been described in the literature.

The results we will elaborate on, are the following:

- Firstly, we show that, using Lagrange multipliers, the least squares optimal model parameters, and the estimated latency and noise sequences, constitute a stationary point of the derivatives of the Lagrangean, which is equivalent to a (potentially large) set of multivariate polynomials. One (or some) of these stationary points provide the global minimum.
- We show that all stationary points of such multivariate polynomial optimization problems, correspond to the eigenvalues of a multi-dimensional (possibly singular) autonomous (nD-)shift invariant dynamical system.
- These eigenvalues can be obtained numerically by applying nD-realization theory in the null space calculated from a so-called Macaulay matrix. This matrix is built from the data, is a highly structured, quasi-Toeplitz matrix and in addition sparse. Revealing the number of stationary points that are finite, their structure and the solutions at infinity, requires a series of rank tests (SVDs) and eigenvalue calculations.
- We show how the parameters of the models are the eigenvalues of a multi-valued eigenvalue problem, and the estimated misfit and latency sequences are (parts) of corresponding eigenvectors. The implication is a serious reduction in computational complexity.
- Finally, we show how also this multi-valued eigenvalue problem ultimately reduces to an 'ordinary' (generalized) eigenvalue problem (i.e. in one variable, which is e.g. the minimal value of the least squares objective function).

Our results put many identification methods that have been described in the signal processing and identification literature, in a new perspective (think of VARPRO, IQML, 'noisy' realization, PEM (Prediction-Error-Methods), Riemannian SVD, ….): all of them generate (often very good) 'heuristic' algorithms, but here for the first time, we will show how all of them in fact are heuristic attempts to find the minimal eigenvalue of a large matrix.

The algorithms we come up with are guaranteed to find the global minima, using the well understood machinery of eigenvalue solvers and singular value decomposition algorithms. We will elaborate on the computational challenges (e.g. the fact that we only need to compute one eigenvalue-eigenvector pair of a large matrix, namely the one corresponding to the global optimum).

We will also elaborate on some preliminary system theoretic interpretations of these results (e.g. that misfit and latency error sequences are also highly structured in a system theoretic sense), and on some ideas on how to incorporate a priori information (e.g. power spectra) into these methods.