FAST SYSTEM IDENTIFICATION AND MODEL
REDUCTION SOLVERS

Vasile Sima ∗ Peter Benner ∗∗

∗ National Institute for Research & Development in Informatics
Bd. Mareșal Averescu, Nr. 8–10, 011455 Bucharest 1, Romania
Fax: (+40)-21-316-1030
E-mail: vsima@ici.ro

∗∗ Fakultät für Mathematik, TU Chemnitz
D-09107 Chemnitz, Germany
Fax : (+49)-371-531-22509
E-mail: benner@mathematik.tu-chemnitz.de

Abstract: Efficient numerical techniques for multivariable system identification and model
reduction are investigated. The techniques are implemented in the system identification
and model reduction toolboxes based on the Fortran 77 Subroutine Library in Control
Theory (SLICOT). Besides highly performant Fortran drivers and computational routines,
these toolboxes provide MATLAB or Scilab interfaces, implementing several algorithmic
approaches. Extensive numerical testing and comparisons with similar MATLAB tools
show that the solvers in these toolboxes are reliable, efficient, and able to solve industrially
relevant problems. Copyright ©2007 IFAC

Keywords: Identification Algorithms, Linear Multivariable Systems, Model Reduction,
Numerical Algorithms, Parameter Estimation, Singular Value Decomposition

1. INTRODUCTION

The availability of powerful system identification and
model reduction tools is very important in practice,
since modern control techniques are increasingly de-
pendent on suitable dynamical models. The numerical
procedures need to be based on reliable and robust
numerical software provided by well-tested and user-friendly software libraries. This paper summa-
izes the functional and computational performances
of the MATLAB 1 identification and model reduction
toolboxes based on the SLICOT Library (Benner
et al., 1999). 2

The SLICOT-based MATLAB toolboxes provide a
user-friendly interface to the highly efficient, robust,
and portable Fortran 77 SLICOT routines. The
MATLAB M-functions contained in these toolboxes
are based on MEX-files calling the Fortran routines.
While the M-functions are destined to all user cat-
egories, the more sophisticated and flexible MEX-
functions are intended for expert users and software
developers. Executable SLICOT MEX-files are pro-
vided for MATLAB running under WINDOWS 95–XP,
Sun Solaris, and Linux.

Most of the SLICOT functionality is concerned with
linear time-invariant (LTI) systems in state-space
form:

\[
\begin{align*}
\dot{x}(t) &= Ax(t) + Bu(t) \\
y(t) &= Cx(t) + Du(t),
\end{align*}
\]

in the continuous-time case, and

\[
\begin{align*}
x_{k+1} &= Ax_k + Bu_k \\
y_k &= Cx_k + Du_k,
\end{align*}
\]

1 MATLAB is a registered trademark of The MathWorks.
2 The SLICOT (Subroutine Library in Control Theory) software
library and the related CACSD tools based on SLICOT were de-
veloped within the Numerics in Control Network (NICONET)
funded by the European Community BRIT-EURAM III RTD The-
matic Networks Programme, see http://www.icm.tu-bs.
de/NICONET. SLICOT can be used free of charge by academic
users, see http://www.slicot.org.
in the discrete-time case, where \( A \) is \( n \times n \), \( B \) is \( n \times m \), \( C \) is \( \ell \times n \), and \( D \) is \( \ell \times m \). All matrices are assumed to be real.

Alternatively, an LTI system (1) or (2) may be represented by a rational transfer-function matrix (TFM),

\[
G(\lambda) = C(\lambda I - A)^{-1}B + D,
\]

where \( \lambda \) is the variable appearing in the Laplace transform, or the \( z \) variable appearing in the \( z \)-transform, respectively.

Sections 2 and 3 provide a short description of the two SLICOT toolboxes emphasizing their potential in solving computationally challenging problems.

2. THE SYSTEM IDENTIFICATION TOOLBOX

Consider a discrete-time system in innovation form,

\[
x_{k+1} = Ax_k + Bu_k + Ke_k,
\]

\[
y_k = Cx_k + Du_k + e_k,
\]

where \( x_k \) is the state vector at time \( k \), \( u_k \) and \( y_k \) are the input and output vectors, respectively, \( \{e_k\} \) is a zero mean white noise sequence, uncorrelated with \( \{u_k\} \) and with the initial state of the system, and \( K \) is the Kalman gain matrix; \( (A, C) \) is assumed observable.

In system identification problems, the system order, \( n \), and the quadruple of system matrices \( (A, B, C, D) \) have to be determined (up to a system similarity transformation) using the input and output data sequences, \( \{u_k\} \) and \( \{y_k\} \). The Kalman gain matrix \( K \) in (4) has often to be found.

The SLICOT calculations are based on subspace identification algorithms, which use a matrix \( H \), built from block-Hankel matrices, using part of the available input-output (I/O) data. Two approaches, MOESP (Multivariable Output Error state-space) (Verhaegen and Dewilde, 1992), and N4SID (Numerical algorithm for Subspace State Space System Identification) (Van Overschee and De Moor, 1994), and their combination are used. The matrix \( H \) is given by

\[
H = \begin{bmatrix}
U_{q+1,q}^{T} & U_{q+1,q,N}^{T} & U_{1,q,r}^{T} & U_{1,r+1,N}^{T} \\
U_{q+1,q}^{T} & U_{q+1,q,N}^{T} & U_{1,q,r}^{T} & U_{1,r+1,N}^{T}
\end{bmatrix},
\]

(4)

where \( H \in \mathbb{R}^{N \times (mq + \ell r + (m + \ell)s)} \), with \( N \leq \bar{N} - s - \min(q, r) + 1 \) (usually \( N \gg mq + \ell r + (m + \ell)s \)),

\[
U_{a,b,c} = \begin{bmatrix}
u_{a} & u_{a+1} & \cdots & u_{a+b-2} & u_{a+b-1} \\
u_{a+1} & u_{a+2} & \cdots & u_{a+b-1} & \cdots & u_{a+c-1} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \ddots \\
u_{b} & u_{b+1} & \cdots & u_{b+c-2} & \cdots & u_{b+c-a-1}
\end{bmatrix},
\]

(5)

and similarly for \( Y_{a,b,c} \).

Where \( s \) denotes the “number of block rows”, and \( s \) usually satisfies \( s \geq n \).

The first step in a subspace identification procedure is to perform a “data compression” by computing the upper triangular factor \( R \) of a QR factorization of the matrix \( H \), \( H = QR \); the matrix \( Q \in \mathbb{R}^{N \times (2m + \ell + s)} \), satisfying \( Q^{T}Q = I_{2m + \ell + s} \), is not needed subsequently. Parts of \( R \) are further used to estimate \( n \) and system matrices. As many theoretical results concerning, e.g., consistency and normality of the estimates of the system matrices hold asymptotically, the algorithms should work with large amounts of data, if available. The standard QR factorization algorithm could be too costly, because \( H \) can have very large dimensions. Since matrix \( H \) is highly structured, efficient data processing is possible by using the problem structure. Two such techniques implemented in the SLICOT Library are based on fast Cholesky and fast QR factorization algorithms, which exploit the special structure of the matrix \( H \).

The Cholesky factorization algorithm efficiently builds the inter-correlation matrix \( W = H^{T}H \), and then factors \( W \), assuming it is positive definite. (In the rare case when this algorithm fails, the usual QR factorization is automatically used.) For the standard N4SID approach, the block-Hankel matrices corresponding to the inputs and outputs are \( H_{1} = U_{1,2r,N}^{T} \) and \( H_{2} = Y_{2,2,N}^{T} \), and \( H = [H_{1} \ H_{2}] \). The definitions extend to multiple I/O data batches. Actually, assuming that the latest batch to be processed is defined by \( H \), then \( W = W + H^{T}H \), where \( W \) corresponds to the already processed data batches. Clearly, \( W = 0 \) if the first (or single) data batch is processed. Two lemmas in (Simar et al., 2004; Sima, 2004) give cheap formulas for computing the symmetric block matrix \( W \) for the N4SID and MOESP approaches, respectively. In the later case, the matrix for the N4SID technique is built, and then it is transformed for the MOESP case.

Consider now the fast QR factorization algorithm for the N4SID approach. Define the shift matrix \( Z = \text{diag}(Z_{2q}, Z_{1r}) \), where \( Z_{2q} \) (\( Z_{1r} \)) is a \( 2q \times 2q \) (\( 2r \times 2r \)) block matrix with \( m \times m \) (\( \ell \times \ell \)) blocks, all zero except for identity blocks on the superdiagonal. A lemma in (Simar et al., 2004) shows that the symmetric matrix \( VW = W - Z^{T}WZ \), called the displacement of \( W \) (Kailath and Sayed, 1995), has a low rank factorization. Specifically, \( VW \) can be written as

\[
VW = G^{T}G, \quad \Sigma = \text{diag}(I_{p}, -I_{q}),
\]

(6)

where \( p = q = m + \ell + 1 \), hence \( VW \) has the rank \( 2(m + \ell + 1) \) at most. The matrix \( G \in \mathbb{R}^{2(m + \ell + 1) \times 2(m + \ell + 1)} \) is called the generator of \( W \), and efficient techniques are available to build the generators and the factor \( R \) (Kailath and Sayed, 1995; Mastronardi et al., 2001).

In practical calculations, the generalized Schur algorithm is used to find all rows of the Cholesky factor of \( W \), \( W = W + H^{T}H \). Householder transformations and hyperbolic rotations are involved. Details are
given in (Mastronardi et al., 2001). For the MOESP approach, the same algorithm as for N4SID is used, but the first two block columns of the resulting upper triangular factor $R$ are interchanged, and then retriangularized, exploiting the structure.

SLICOT System Identification Toolbox can also be used to identify Wiener systems. A discrete-time Wiener system has a state-space representation

$$
x_{k+1} = Ax_k + Bu_k, \quad z_k = Cx_k + Du_k,
$$

where $f(\cdot)$ is a nonlinear vector function from $\mathbb{R}^\ell$ to $\mathbb{R}^\ell$. Briefly speaking, a Wiener system consists of a linear dynamic block followed by a static nonlinearity. The system identification problem for (7) requires to find an approximation of $f(\cdot)$, besides estimating $n$, the quaduple $(A, B, C, D)$, and the initial conditions. The implemented approach uses a state-space representation for the linear part and a single layer neural network to model the static nonlinearity. Fast subspace identification algorithms are employed for estimating the linear part, based on the available input-output data. The resulted state-space model is used for finding an approximate model of the nonlinear part by a Levenberg-Marquardt (LM) algorithm. Finally, the whole model is refined using a specialized, structure-exploiting LAPACK-based scaling-invariant LM algorithm.

The output normal form is used to parameterize the linear part. The parameters corresponding to the nonlinear part come first in the global parameter vector. Using this ordering, the Jacobian matrices in the multi-output case ($\ell > 1$) have a block diagonal form, with an additional block column at the right. This structure is preserved in a QR factorization with column pivoting restricted to each block column. The rank deficient case is also covered. The Jacobian is computed analytically, for the nonlinear part, and numerically, for the linear part. The implementation is memory conserving and significantly faster than standard LM algorithms or specialized LM calculations based on conjugate gradients (without preconditioning) for solving linear systems. Details and numerical results are given, e.g., in (Sima, 2003a; Sima, 2003b).

2.1 Functionality of the Toolbox

Summarizing, SLICOT System Identification Toolbox includes SLICOT-based MATLAB and Fortran tools for linear and Wiener-type, time-invariant discrete-time multivariable systems. The approaches MOESP, N4SID, and their combination, are used to identify linear systems, and to initialize the parameters of the linear part of a Wiener system. The toolbox functionalities include:

- identification of linear discrete-time state-space systems $(A, B, C, D)$;
- identification of linear discrete-time state-space matrices for such systems;
- estimation of the associated Kalman gain matrix;
- estimation of the initial state;
- conversion from/to a state-space representation to/from the output normal form;
- identification of discrete-time Wiener systems;
- computation of the output of Wiener systems.

Attractive features of this toolbox include:

- possible speed-up factors larger then 10 in comparison with the commonly used software tools;
- standard or fast techniques for data compression (exploiting the block-Hankel structure);
- fast estimation of system models of various, specified orders;
- ability to process multiple data batches;
- specialized, structure-exploiting LM algorithm using block QR factorization with pivoting;
- optional assessing the quality of the intermediate results using the associated condition numbers.

2.2 Performance Results

Performance evaluation of the system identification software has been performed using data sets from the DAISY collection, publicly available from the Internet site www.esat.kuleuven.be/sista/daisy. Accuracy and efficiency comparisons of the SLICOT linear systems identification software and the available subspace-based codes for 25 applications are presented in (Sima et al., 2004). New results are given below, comparing the SLICOT code slmoen4 and MATLAB function n4sid (Ljung, 2000). The main SLICOT-based call for standard calculations was

$$[\text{sys}, \text{K}, \text{rcnd}] = \text{slmoen4}(s, y, u, n, \text{alg});$$

The results have been obtained on an Intel Pentium 4 computer, at 3 GHz, with 1 GB RAM, under Windows XP, with Compaq Visual Fortran V6.5 and optimized LAPACK and BLAS (available in MATLAB), under MATLAB 7.0.4.365 (R14) Service Pack 2.

Figure 1 shows a timing comparison of slmoen4 with fast QR factorization ($\text{alg} = 2$) vs MATLAB 7.0.4 n4sid with standard QR factorization and default options, but with order $\leq n$, 'N4Weight' = 'MOESP', and four cases (a)-(d) for the possible combinations of values for 'N4H' := 'N4Horizon' and 'Cov' := 'CovarianceMatrix' (see the caption). Here $n$ is the chosen order of the system, and $s$ is the number of block rows, set to the same values as for slmoen4 (see, e.g., (Sima et al., 2004)). Setting 'Cov' = 'None' avoids the calculation of the covariance information, reducing the execution times. Specifically, Figure 1 shows the ratios between the execution times of n4sid and slmoen4—the speed-up factors—for each application and each case (a)-(d). Clearly, the use of a fast algorithm is very advantageous. The figure also illustrates the strong influence...
the default value for the corresponding n4sid parameter. Similarly, ‘Auto’ (as in cases (a) and (b)) is the default value for the parameter ‘N4Horizon’. But for large systems, the use of these default values involves significant computational effort. The largest speed-up factors have been usually obtained in case (a) (and other SLICOT functions) with $\text{alg} = \text{QR}$ factorization and de-

However, for large $\bar{N}$, grows linearly with $\bar{N}$, for large $\bar{N}$.

3. THE MODEL AND CONTROLLER REDUCTION TOOLBOX

SLICOT provides a wide variety of model (order) reduction techniques for LTI systems of the form (1) or (2). Model reduction is often used in system analysis and is mostly absolutely necessary for the computation of controllers. The reason is that modern and robust controllers like those based on LQG, $H_2$, or $H_\infty$ design techniques are themselves LTI systems of the form (1) or (2) with state dimension $N \geq n$, while in practice, only very low values $N$ should be used. The reduction of $n$ can be achieved using model reduction techniques, while $N$ can be directly reduced using controller reduction techniques, see, e.g., (Obinata and Anderson, 2001; Varga, 2001; Varga, 2003). In contrast to the MATLAB Control and Robust Control Toolboxes, SLICOT offers several controller reduction functions. Due to space limitations, here we will consider only continuous-time LTI systems although most SLICOT routines can be applied also to discrete-time systems.

Applying the Laplace transformation to (1) (with $x(0) = 0$), the system dynamics are then given as $Y(s) = G(s)U(s)$, where $U, Y$ are the Laplace transforms of $u, y$, respectively, and $G$ is given in (3).

The aim of model reduction is to find an LTI system,

$$\dot{x}(t) = \hat{A}x(t) + \hat{B}u(t),$$

$$\hat{y}(t) = \hat{C}x(t) + \hat{D}u(t),$$

of order $r, r \ll n$, such that the associated TFM $\hat{G}(s) = \hat{C}(sI - \hat{A})^{-1}\hat{B} + \hat{D}$ approximates the original TFM $G(s)$. This is motivated by the inequality $\|y - \hat{y}\|_2 \leq \|G - \hat{G}\|_\infty \|u\|_2$, where $\| \cdot \|_2$ corresponds to the $L_2$-norm and $\| \cdot \|_\infty$ is the $H_\infty$-norm. Note that model and controller reduction for unstable systems is possible by an additive decomposition of the transfer-function into its stable and unstable parts or by a coprime factorization (where both factors are stable), see, e.g., (Varga, 2003).

Methods that attempt to minimize $\|G - \hat{G}\|$ are called absolute error methods while relative error methods try to minimize $\|\Delta\|$, where $\Delta$ is implicitly defined by $G - \hat{G} = \Delta G$. Nevertheless, balanced truncation and related methods can be used to obtain good approximations using either one of these measures. An alternative is to use the Hankel (semi-)norm of (1) which is defined as the maximum Hankel singular value of $G(s)$ for $s$ on the imaginary axis. Formulas for computing the best Hankel norm approximation to a given stable system $G$ can for instance be found in (Antoulas, 2005; Obinata and Anderson, 2001) and references therein.

There is a vast variety of model reduction techniques serving different purposes; in case of linear systems, it seems that modal truncation and the related techniques of substructuring and static condensation, Padé and Padé-type approximations, and balancing-related truncation techniques play the most prominent role; see the recent monographs and surveys (Antoulas, 2005; Benner et al., 2005; Benner et al., 2006; Obinata and Anderson, 2001). SLICOT’s model and controller reduction routines are all based on the latter approach. One reason is that SLICOT uses dense linear algebra while modal as well as Padé techniques have only advantages if large and sparse systems have to be reduced and $n$ is too large to use dense linear algebra—regarding their model reduction abilities, they are quite inferior to balancing-related techniques.

Balanced truncation is based on finding a state-space transformation which balances the controllability Gramian $P$ versus the observability Gramian $Q$ of

Fig. 1. slmoen4 with fast QR versus MATLAB 7.0.4 n4sid with QR factorization and default options, but $\text{order} = n$, ‘N4Weight’ = ‘MOESP’, and: (a) ‘Cov’ = [ ], ‘N4H’ = ‘Auto’; (b) ‘Cov’ = ‘None’, ‘N4H’ = ‘Auto’; (c) ‘Cov’ = [ ], ‘N4H’ = [ss]; (d) ‘Cov’ = ‘None’, ‘N4H’ = [ss].

the n4sid options can have on the execution times. It should be mentioned that n4sid computes the co-variance matrices of the estimated systems in the cases (a) and (c), while SLICOT function does not compute them. These results are shown since ‘Cov’ = [ ] is the default value for the corresponding n4sid parameter. Similarly, ‘Auto’ (as in cases (a) and (b)) is the default value for the parameter ‘N4Horizon’. But for large systems, the use of these default values involves significant computational effort. The largest speed-up factors have been usually obtained in case (a), normally followed by (c). Even if ‘Cov’ = ‘None’ and ‘N4H’ = [ss], as for slmoen4), the SLICOT function can be much more efficient than n4sid, see the case (b) (and the case (d)) in Fig. 1. Moreover, the timing results for slmoen4 (and other SLICOT functions) with $\text{alg} = 1$ or $\text{alg} = 2$ are essentially independent on the number of data samples $\bar{N}$, used for identification (see (Sima, 2004)). This is due to the use of fast factorization algorithms. On the other hand, the computing time for n4sid, which uses standard QR factorization of the block-Hankel-block matrix $\hat{H}$, grows linearly with $\bar{N}$, for large $\bar{N}$.
the system (1). The Gramians are given as the solutions of the Lyapunov equations
\[
AP + PA^T + BB^T = 0, \quad A^TQ + QA + C^T C = 0.
\] (9)

A minimal and stable LTI system can be transformed by a state-space transformation (change of coordinates in state-space) such that the Gramians \( P, Q \) become equal and diagonal and the diagonal elements \( \sigma_j, j = 1, \ldots, n \), are monotonically decreasing positive real numbers, called the Hankel singular values of the LTI system (1). The reduced-order model is obtained by truncating the balanced state-space representation of the system at an order \( r \) so that \( \sigma_r > \sigma_{r+1} \). The so-obtained reduced-order model is stable and satisfies the error bound
\[
\sigma_{r+1} \leq \|G - \hat{G}\|_\infty \leq 2 \sum_{k=r+1}^{n} \sigma_k,
\] (10)

which allows an adaptive choice of \( r \) given an error tolerance. In the SLICOT toolbox, there are also several functions implementing balancing-related methods that can be used if other system properties are to be preserved (e.g., minimum-phase) or if controller reduction is desired. To overcome a disadvantage of balanced truncation of not preserving the steady-state performance, i.e., \( G(0) \neq \hat{G}(0) \), it can be combined with singular perturbation approximation (SPA).

### 3.1 Functionality of the Toolbox

The SLICOT Model and Controller Reduction Toolbox for MATLAB provides numerically reliable and efficient functions for balanced truncation, singular perturbation approximation, balanced stochastic truncation, frequency-weighting balancing, Hankel-norm approximation, coprime factorization, etc. The algorithms are based on methods with theoretically sound mathematical background described well in (Antoulas, 2005; Obinata and Anderson, 2001). The functionality of the toolbox includes

- order reduction for continuous-time and discrete-time LTI systems and controllers;
- order reduction for stable or unstable models/controllers;
- additive and relative error model reduction;
- frequency-weighted reduction with special stability/performance enforcing weights;
- coprime factorization-based reduction of state feedback and observer-based controllers.

The main features of the toolbox are:

- computational reliability using square-root and balancing-free accuracy enhancing techniques;
- high numerical efficiency, using latest developments and structure-exploiting algorithms;
- enhanced functionality, e.g. controller reduction.

For a more detailed description of the SLICOT Fortran 77 routines for model and controller reduction see (Varga, 2003).

![Fig. 2. Comparison of MATLAB functions for singular perturbation approximation for SISO (top) and MIMO (bottom, \( m = 5, \ell = 10 \)) systems.](image)
4. CONCLUSIONS

The results show that the fast and reliable system identification and model/controller reduction algorithms and solvers included in the SLICOT toolboxes can be safely used, and they are significantly more efficient than the existing MATLAB codes.

REFERENCES


