SLICOT Working Note 2002-1

Enhanced Services for Remote Model Reduction of Large-Scale Dense Linear Systems[1]

Peter Benner[2], Rafael Mayo[3], Enrique S. Quintana-Ortí[4], and Gregorio Quintana-Ortí[5]

January 2002

[1]This document presents research results of the European Community BRITE-EURAM III Thematic Networks Programme NICONET (contract number BRRT-CT97-5040) and is distributed by the Working Group on Software WGS. *WGS secretariat*: Mrs. Ida Tassens, ESAT - Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001-Leuven-Heverlee, BELGIUM. This report is available by anonymous ftp from `wgs.esat.kuleuven.ac.be/pub/WGS/REPORTS/SLWN2002-1.ps.Z`
[2]Arbeitsbereich Mathematik, Technische Universität Hamburg-Harburg, D-21071 Hamburg, Germany
[3]Depto. de Ingeniería y Ciencia de Computadores, Universidad Jaume I, 12080–Castellón, Spain.
[4]Same address as second author.
[5]Same address as second author.

Abstract

This paper describes enhanced services for remote model reduction of large-scale, dense linear time-invariant systems. Specifically, we describe a Web service and a Mail service for model reduction on a cluster of Intel Pentium-II architectures using absolute error methods. Experimental results show the appeal and accessibility provided by these services.

Contents

1 Introduction

The development of the Internet and its popularization via the world wide web (www) and electronic mail (e-mail) is producing a tremendous impact in our society. Nowadays, many people employ e-mail to communicate daily and use the www to access a large number of services which is increased every day. The extension of these services has recently reached the area of scientific computing in projects like NetSolve [1], LAPK (`http://lapk3.hsc.usc.edu/lapk`), the SLICOT Web Environment [10], etc.

In this paper we further explore both the www and e-mail as a means of offering enhanced services to a wide range of the scientific community who do not have experience with parallel computing and/or lack of the access to a parallel architecture. In particular, we present prototypes of Web and Mail services for parallel model reduction of linear time-invariant (LTI) systems.

Specifically, the problem considered here can be stated as follows. Consider the transfer function matrix (TFM) $G(s) = C(sI - A)^{-1}B + D$, and the associated stable, but not necessarily minimal, realization of an LTI system,

$$\begin{array}{rcll} \dot{x}(t) & = & Ax(t) + Bu(t), & t > 0, \qquad x(0) = x^0, \\ y(t) & = & Cx(t) + Du(t), & t \geq 0, \end{array} \tag{1}$$

where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, and $D \in \mathbb{R}^{p \times m}$; the number of states, n, is also said to be the order of the system. In model reduction, the goal is to find a reduced-order LTI system

$$\begin{array}{rcll} \dot{\hat{x}}(t) & = & \hat{A}\hat{x}(t) + \hat{B}\hat{u}(t), & t > 0 \qquad \hat{x}(0) = \hat{x}^0, \\ \hat{y}(t) & = & \hat{C}\hat{x}(t) + \hat{D}\hat{u}(t), & t \geq 0, \end{array} \tag{2}$$

of order r, $r \ll n$, and associated TFM $\hat{G}(s) = \hat{C}(sI - \hat{A})^{-1}\hat{B} + \hat{D}$ which approximates $G(s)$.

Model reduction of large-scale systems arises, among others, in control of large flexible structures or large power systems, circuit simulation, VLSI design, etc. [7, 8, 11, 15]. Reliable methods for model reduction usually require $\mathcal{O}(n^3)$ flops and storage for $\mathcal{O}(n^2)$ numbers. While systems of order n in the hundreds may be treated on current desktop computers (using libraries like SLICOT[1]), large-scale applications clearly require the use of parallel computing techniques to obtain the reduced system.

The paper is structured as follows. In Section 2 we briefly revisit the methods offered by our parallel library. Section 3 describes the Web and Mail services provided to facilitate the use of the parallel kernels. Implementation details of these services and the integration are given in Section 4. Finally, concluding remarks follow in Section 5.

2 PLiCMR: A Parallel Library for Model Reduction

Our parallel model reduction library provides absolute error methods, based on similarity transformations, which try to minimize $\|\Delta_a\|_\infty = \|G - \hat{G}\|_\infty$. Here, $\|G\|_\infty$ denotes the $\mathcal{L}_\infty$- or $\mathcal{H}_\infty$-norm of a stable, rational matrix function defined as

$$\|G\|_\infty = \operatorname*{ess\,sup}_{\omega \in \mathbb{R}} \sigma_{\max}(G(\jmath\omega)), \tag{3}$$

where $\jmath := \sqrt{-1}$ and $\sigma_{\max}(M)$ is the largest singular value of the matrix M. Specifically, three different methods are offered as part of the library: Balanced truncation (BT) methods [14, 16, 17, 18], singular perturbation approximation (SPA) methods [13], and Hankel-norm approximation (HNA) methods [12].

The parallel kernels in PLiCMR follow closely analogous routines from SLICOT. The parallelization heavily relies on the use of the parallel infrastructure in ScaLAPACK (including PBLAS and BLACS),

[1]Available from `http://www.win.tue.nl/niconet/NIC2/slicot.html`.

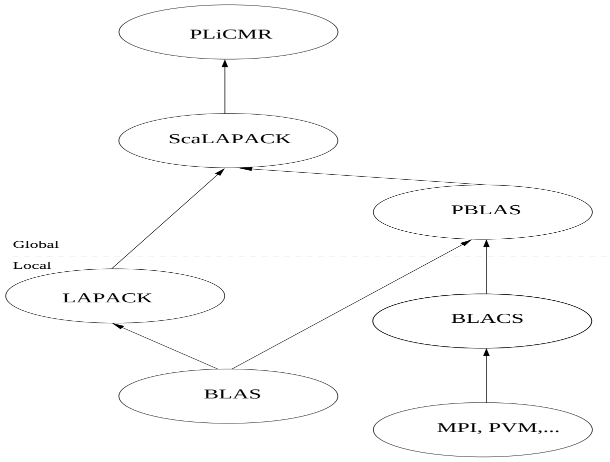

Figure 1: PLiCMR and the underlying libraries.

and the serial computational libraries LAPACK and BLAS (see Figure 1). Therefore, the library can be migrated to any parallel architecture were ScaLAPACK, LAPACK, BLAS, and a communication library like MPI or PVM is available. Details of the parallelization are given in [2, 3, 4, 5].

The calling sequence of the parallel routines follow the SLICOT style. Thus, subroutines `PAB09AX` for BT model reduction, `PAB09BX` for SPA model reduction, and `PAB09CX` for HNA model reduction present the following interfaces:

```
SUBROUTINE PAB09AX( DICO, JOB, EQUIL, ORDSEL, N, M, P, MAXIT, NR,
                    A, DESCA, B, DESCB, C, DESCC, HSV, TOL1,
                    IWORK, LIWORK, DWORK, LDWORK, INFO )

SUBROUTINE PAB09BX( DICO, JOB, EQUIL, ORDSEL, N, M, P, MAXIT, NR,
                    A, DESCA, B, DESCB, C, DESCC, D, DESCD, HSV,
                    TOL1, TOL2, IWORK, LIWORK, DWORK, LDWORK, INFO )

SUBROUTINE PAB09CX( DICO, EQUIL, ORDSEL, N, M, P, MAXIT, NR,
                    A, DESCA, B, DESCB, C, DESCC, D, DESCD, HSV,
                    TOL1, TOL2, IWORK, LIWORK, DWORK, LDWORK, INFO )
```

The parameters of the routines can be grouped as follows:

- `DICO, JOB, EQUIL, ORDSEL, MAXIT`: Input parameters specifying, respectively, the type of system (discrete or continuous), job to perform (square-root or balancing-free square-root), whether to equilibrate the system, the method to select the order of the reduced system, and the maximum number of iterations for the iterative solvers employed to compute the Gramians of the system.
- `N, M, P`: Dimensions of the LTI system.
- `NR`: Desired order of the resulting system.
- `A, DESCA, B, DESCB, C, DESCC` (and `D, DESCD`): Matrices of the system and their corresponding descriptors (see ScaLAPACK [6]).
- `HSV`: Computed Hankel singular values of the system.
- `TOL1` (and `TOL2`): Tolerances to determine the order of the reduced system.

– IWORK, LIWORK, DWORK, LDWORK: Workspaces and dimensions.

– INFO: Return state.

More details on the parallel library can be obtained at http://spine.act.uji.es/~plicmr.

3 Enhanced Services for Remote Model Reduction

Our Web and Mail services are provided on a parallel cluster[2], composed of 32 Intel Pentium-II personal computers, running a Linux Operating System, and connected with a Gigabit Myrinet network. With this architecture, model reduction of dense LTI systems with up to 5000 states is possible. Both services can be accessed at http://spine.act.uji.es/~plicmr.

3.1 A Web service for remote access

The Web service for parallel model reduction is accessed from any standard web browser, and does not need the installation of any software on the local computer. Figure 2 shows a snapshot of the interface the user is offered when he/she enters the Web service. The interface describes briefly the service that is offered, and allows the user to register with the Web server, submit a job, check the status of submitted jobs, and study the format of the output of the job execution.

Access to the service is validated during job submission with a login (user identifier) and a password (both provided on registration). Job submission is performed via a web interface where the user can specify the method to employ (BT, SPA, or HNA), and the parameters of the parallel routine DICO, JOB, EQUIL, ORDSEL, N, M, P, NR, TOL1 (and TOL2); the interface also allows the user to choose the number of processors to use in order to execute the job, the format of the files containing the matrices (compressed or not, and the compression tool), and an e-mail address where the results of the execution (the reduced system) should be returned. An interactive response via the web interface is probably not desirable here as the response time depends on the system load and, even in the case of lightly-loaded systems, a large-scale problem may require a large execution time. Finally, the data matrices can be uploaded from the user local file system. A snapshot of this interface is shown in Figure 3.

The data matrices can be provided in separate ASCII files (one file per matrix), using dense or sparse formats. Dense matrices must be stored in the file in column-major order, while sparse matrices are stored in coordinated storage format, i.e., with tuples of the form (row, column, value). Two characters in the first two lines of the file specify the matrix in the file (A, B, C, or D), and the format (sparse or dense). Table 1, e.g., shows the contents of three files which store a 7×7 matrix A with 16 nonzero entries in sparse format, a 7×2 matrix B in dense format, and a 3×7 matrix C also in dense format.

As an example, if the user submits a job with the parameters specified in Figure 3, and the data matrices in table 1, an e-mail is returned to user1@spine.act.uji.es with the contents in Table 2. The reduced model is returned as three different files attached to the e-mail, with the contents shown in Table 3. The matrix files are processed using the same compression tool the user employed to provide the data matrices. The Hankel singular values of the system are also provided in an attached file in order of decreasing magnitude.

[2]The cluster is owned by the Parallel Scientific Research Group, at the University Jaume I; http://spine.act.uji.es/psc.html.

File for A	File for B	File for C
'A'	'B'	'C'
'S'	'D'	'D'
1 1 -0.04165	0.00000	1
2 1 -5.21000	12.50000	0
4 1 0.54500	0.00000	0
2 2 -12.5000	0.00000	0
3 2 3.33000	0.00000	0
1 3 4.92000	0.00000	0
3 3 -3.33000	0.00000	0
1 4 -4.92000	0.00000	0
5 4 4.92000	0.00000	0
4 5 -0.54500	0.00000	0
5 5 -0.04165	0.00000	0
6 5 -5.21000	0.00000	1
6 6 -12.5000	0.00000	0
7 6 3.33000	0.00000	0
5 7 4.92000	12.50000	0
7 7 -3.33000	0.00000	1
		0
		0
		0
		0
		0
		0

Table 1: Contents of data files for the original system.

```
1.  JOB ID. = 00001
   (METHOD='AX', DICO='C', JOB='B', EQUIL='N', ORDSEL='F',
    N=7, M=2, P=3, NR=3,
    TOL1=1.0E-14, TOL2=0.0,
    NP=4, COMPRESS='GZ').
2.  NR = 3
3.  NP = 4
4.  TIME = 0.3507E+1
5.  INFO = 0
```

Table 2: Contents of an e-mail with the results of the execution.

File for reduced A	File for reduced B	File for reduced C
'A'	'B'	'C'
'S'	'D'	'D'
1 1 -0.3967426451105343E+00	-0.9986807045917956E+00	-0.9921538597682402E+00
2 1 0.3087135117686742E+01	0.7253607356343710E+00	0.1612071231115642E+00
3 1 0.2069870357469632E-13	0.1059129453177785E+01	0.9921538597682398E+00
1 2 -0.3027914808756307E+01	0.9986807045917960E+00	-0.6334895014421159E+00
2 2 -0.2524022142684217E+00	-0.7253607356343844E+00	-0.4996783932944683E+00
3 2 0.2036863516394205E-14	0.1059129453177775E+01	0.6334895014421047E+00
1 3 -0.1833000304275398E-13		0.1059129453177776E+01
2 3 0.4143105807432115E-14		-0.2805932011253109E-14
3 3 -0.5849191362119688E+00		0.1059129453177784E+01

Table 3: Contents of data files for the reduced order system.

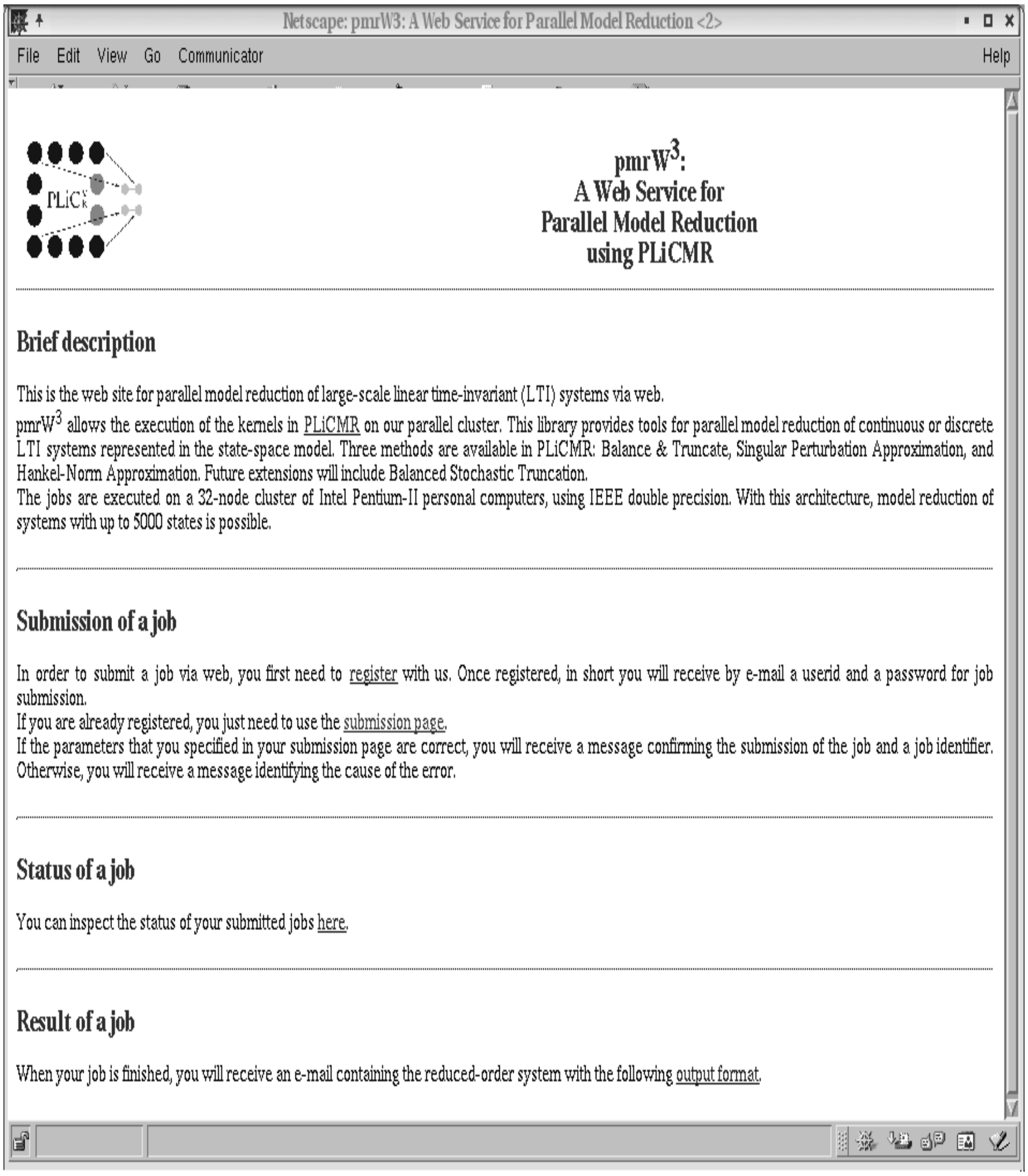

Figure 2: Snapshot of the interface for the Web service.

3.2 A Mail service for remote access

The Mail server for parallel model reduction accepts jobs submitted with any standard e-mail application. Figure 4 shows a snapshot of the interface the user is offered when he/she enters the Mail service. The interface describes briefly the service that is offered, allows the user to register with the Mail server, and provides information on the format of the submission e-mail, the status of submitted jobs, and the format of the output of the job execution.

As could be expected, the contents of an e-mail submitting a job are exactly the same as those which have to be specified for the Web service. Table 4 shows the contents of an e-mail submitting a job as the

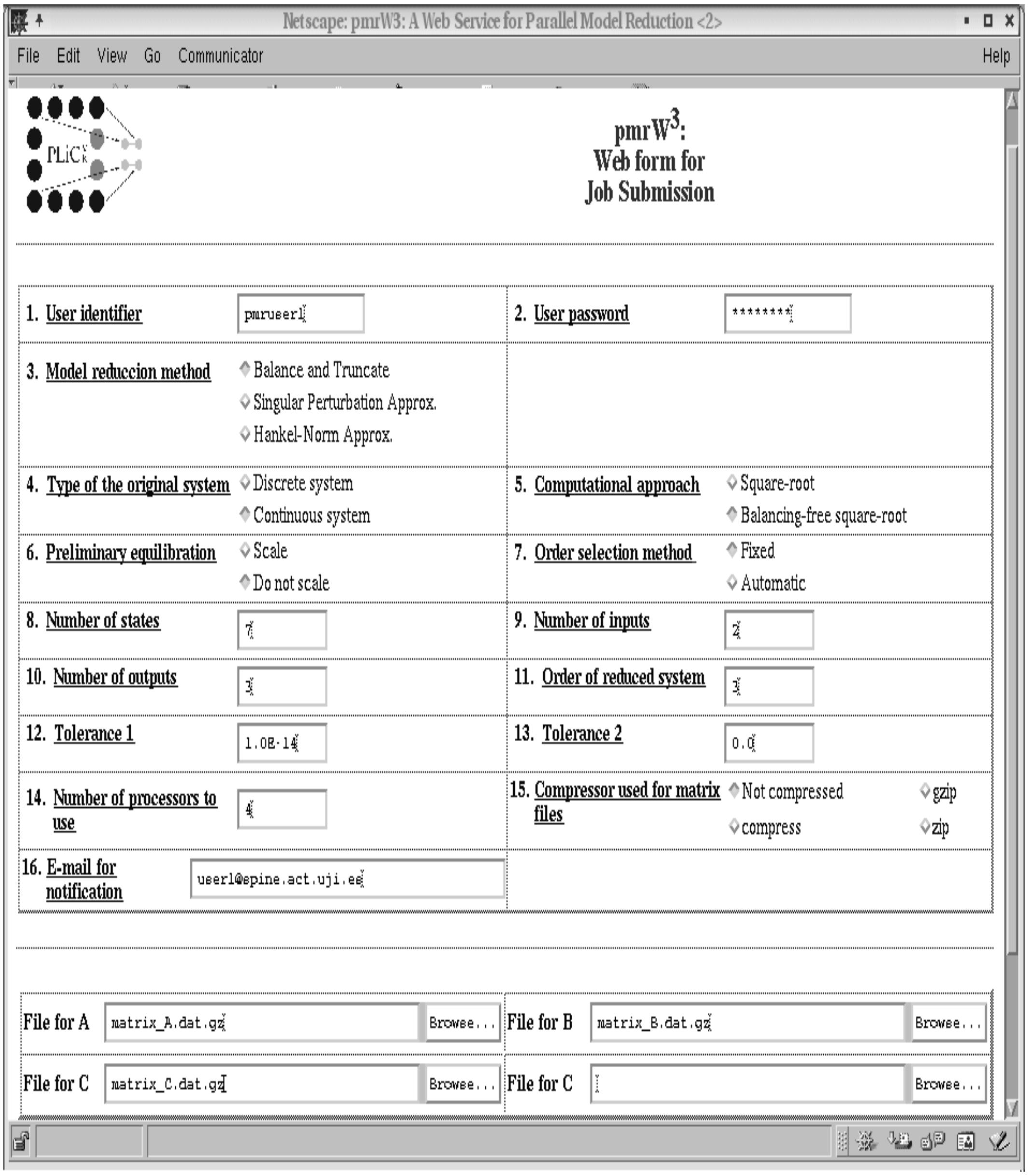

Figure 3: Snapshot of the interface for job submission.

one specified in Figure 3.

The data matrices have to be attached as separate files following the same format described in the previous subsection. After execution, the result (reduced order system) is returned in an e-mail with the same structure as that described for the Web service. Tables 2 and 3 show the results returned to the user after submitting the job specified in the e-mail in Table 4, with the data matrices in Table 1.

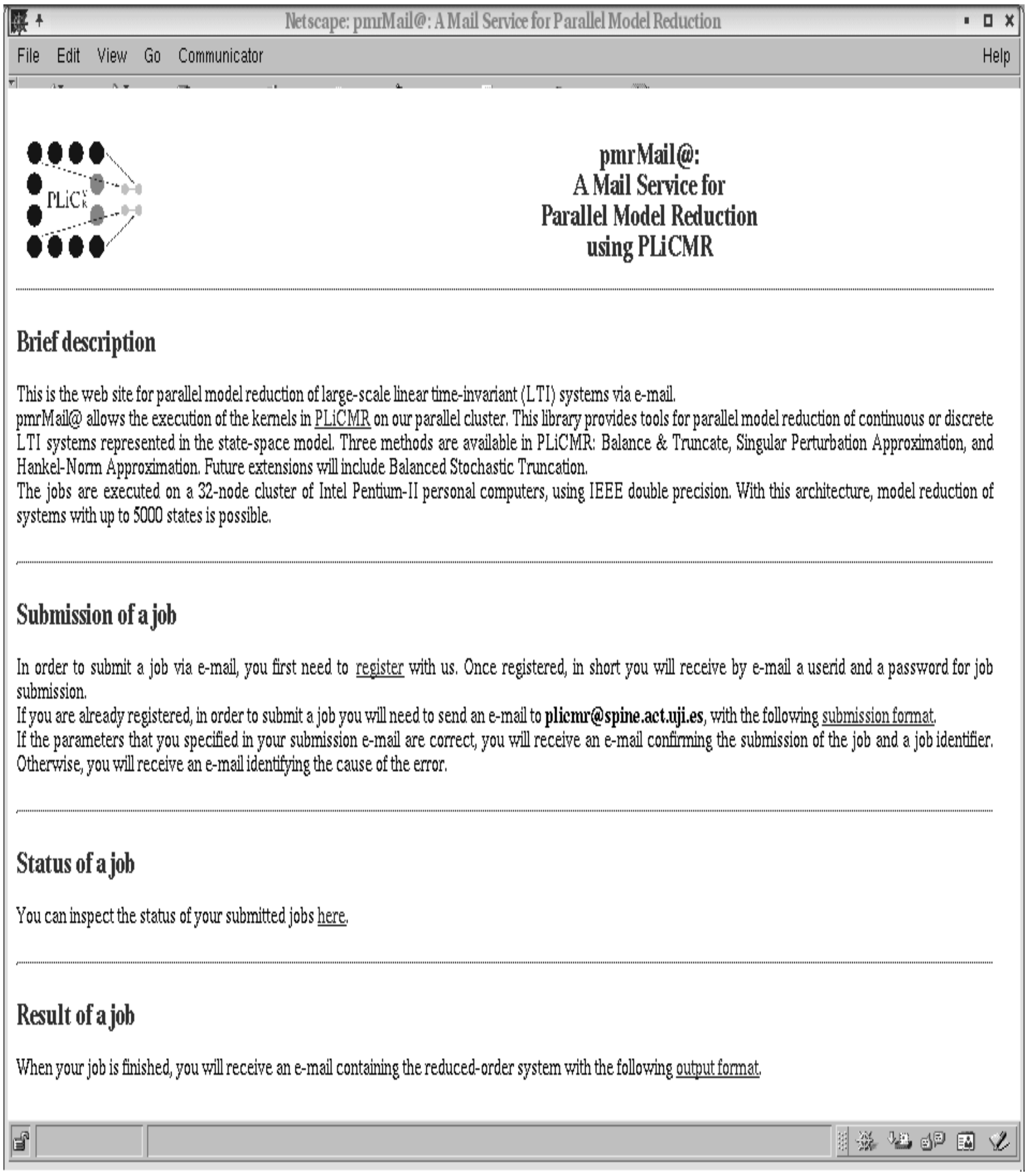

Figure 4: Snapshot of the interface for the Mail service.

4 Implementation Details

The Web and Mail services are structured as three different components (see Figure 5):

- The Web Queue System Interface (WQSI), which assists the Web Server in order to identify the requested service and prepares the necessary data for the job. This interface serves as an intermediary between the user requests and the Queue System, so that the implementation of the Queue System is completely independent of the method used to submit the request (web interface or e-mail). Each new service request is preprocessed by the WQSI, which verifies the format of the request,

'pmruser1'
'mypasswd'
'AX'
'C'
'B'
'N'
'F'
7
2
3
3
1.0E-14
0.0
4
'GZ'
'user1@spine.act.uji.es'

Table 4: Contents of an e-mail submitting a job to the Mail service.

generates the necessary files, and uses the Queue System Client in order to submit a new job to the parallel system.

- The Mail Queue System Interface (MQSI), which assists the Mail Server, with a functionality similar to that of the WQSI.
- The Queue System which, using the input preprocessed by the WQSI or the MQSI, performs the job and returns the result to the remote user. The Queue System Server controls the execution of jobs on the cluster. This component maintains a list of idle processors and schedules the jobs using a FIFO policy enhanced with priorities.
  The Queue System Client is used to start the execution of a new job. This component opens a connection with the Queue System Server and, using a predefined communication protocol, submits a new job.

5 Concluding Remarks

In this paper we have presented enhanced services for remote model reduction of LTI systems via web or e-mail. These environments can be accessed from any standard web browser or e-mail application and provide an easy-to-use and powerful testing tool for researchers dealing with model reduction of large-scale LTI systems.

References

[1] D. Arnold, S. Agrawal, L.S. Blackford, J. Dongarra, M. Miller, S. Vahdiyar, K. Sagi, and Z. Shi. Users's guide to NetSolve. Technical Report UT-CS-01-467, University of Tennessee at Knoxville, Computer Science Dept., 2001. Available from `http://icl.cs.utk.edu/netsolve`.

[2] P. Benner, E.S. Quintana-Ortí, and G. Quintana-Ortí. Balanced truncation model reduction of large-scale dense systems on parallel computers. Mathematical and Computer Modeling of Dynamical Systems, Vol. 6(4), pp. 383–405, 2000.

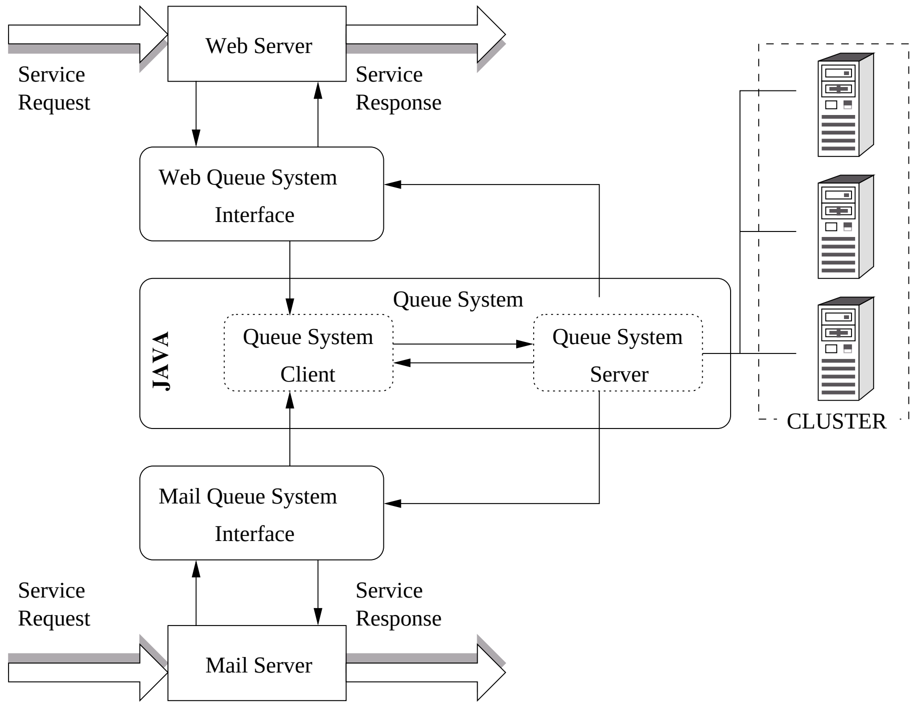

Figure 5: Components of the Web and Mail service system.

[3] P. Benner, E.S. Quintana-Ortí, and G. Quintana-Ortí. Efficient numerical model reduction methods for discrete-time systems. Proceedings 3rd IMACS Symposium on Mathematical Modeling, volume 1, pages 277–280, 2000.

[4] P. Benner, E.S. Quintana-Ortí, and G. Quintana-Ortí. Singular perturbation approximation of large, dense linear systems. In *Proc. 2000 IEEE Intl. Symp. CACSD, Anchorage, Alaska, USA, September 25–27, 2000*, pages 255–260. Omnipress, Madison, WI, 2000.

[5] P. Benner, E.S. Quintana-Ortí, and G. Quintana-Ortí. PSLICOT routines for model reduction of stable large-scale systems. Proceedings Third NICONET Workshop on Numerical Software in Control Engineering, pages 39–44, 2001.

[6] L.S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R.C. Whaley. *ScaLAPACK Users' Guide*. SIAM, Philadelphia, PA, 1997.

[7] J. Cheng, G. Ianculescu, C.S. Kenney, A.J. Laub, and P. M. Papadopoulos. Control-structure interaction for space station solar dynamic power module. *IEEE Control Systems*, pages 4–13, 1992.

[8] P. Y. Chu, B. Wie, B. Gretz, and C. Plescia. Approach to large space structure control system design using traditional tools. *AIAA J. Guidance, Control, and Dynamics*, 13:874–880, 1990.

[9] U.B. Desai and D. Pal. A transformation approach to stochastic model reduction. *IEEE Trans. Automat. Control*, AC–29:1097–1100, 1984.

[10] E. Elmroth, P. Johansson, B. Kågström, and D. Kressner. A Web computing environment for the SLICOT library. SLICOT Working Note 2001–2, June 2001.

Available from `http://www.win.tue.nl/niconet/NIC2/reports.html`, service accesible at `http://wc2.hpc2n.umu.se`.

[11] L. Fortuna, G. Nummari, and A. Gallo. *Model Order Reduction Techniques with Applications in Electrical Engineering*. Springer-Verlag, 1992.

[12] K. Glover. All optimal Hankel-norm approximations of linear multivariable systems and their L^∞ norms. *Internat. J. Control*, 39:1115–1193, 1984.

[13] Y. Liu and B.D.O. Anderson. Controller reduction via stable factorization and balancing. *Internat. J. Control*, 44:507–531, 1986.

[14] B. C. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Trans. Automat. Control*, AC-26:17–32, 1981.

[15] C.R. Paul. *Analysis of Multiconductor Transmission Lines*. Wiley–Interscience, Singapur, 1994.

[16] M.G. Safonov and R.Y. Chiang. A Schur method for balanced-truncation model reduction. *IEEE Trans. Automat. Control*, AC–34:729–733, 1989.

[17] M.S. Tombs and I. Postlethwaite. Truncated balanced realization of a stable non-minimal state-space system. *Internat. J. Control*, 46(4):1319–1330, 1987.

[18] A. Varga. Efficient minimal realization procedure based on balancing. In *Prepr. of the IMACS Symp. on Modeling and Control of Technological Systems*, volume 2, pages 42–47, 1991.

[19] A. Varga and K. H. Fasol. A new square–root balancing–free stochastic truncation model reduction algorithm. In *Prepr. 12th IFAC World Congress*, volume 7, pages 153–156, Sydney, Australia, 1993.