Fast Time Domain Simulation for Large Order Linear Time-Invariant Systems

Kin Cheong Sou* and Olivier L. de Weck†
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Time domain simulation is an essential technique in multidisciplinary design optimization (MDO). Unfortunately it can be time consuming and cause memory saturation problems when systems get large, sampling rates are high and input time histories are long. In this paper, an efficient simulation scheme is presented to simulate large order continuous-time linear time-invariant (LTI) systems. The $A$ matrix (assumed to be block-diagonalizable) of the system is first diagonalized. Then, subsystems of manageable dimensions and bandwidth are formed and multiple sampling rates can be chosen to associate with the subsystems. Each subsystem is then discretized using a $O(n_s)$ discretization scheme, where $n_s$ is the number of state variables. Next, a sparse matrix recognizable $O(n_s)$ discrete-time system solver (i.e. matrix-vector product solver) is employed to compute the history of the state and the output. Finally, the response of the original system is obtained by superposition and interpolation of the subsystem responses. In practical engineering applications, closing feedback loops, cascading filters (shaping filters or other feedforward compensators) and other structures can hinder the efficient use of the simulation scheme (e.g. by destroying the block diagonal structure of the $A$ matrix). Solutions to these problems are also addressed in the latter part of the paper. The simulation scheme, implemented as a MATLAB script $\text{newlsim.m}$, is compared with the established LTI system simulator $\text{lsim.m}$ in MATLAB and is shown to be superior in the case of medium and large order systems. The 2184 state variable Space Interferometry Mission (SIM v2.2) simulation is enabled with the proposed technique (standard simulation fails due to excessive memory requirements) and a computer time savings factor of $\approx 50$ is demonstrated without loss of accuracy. Aside from handling realistic applications the simulator brings time domain simulation on par with frequency domain and Lyapunov analyses, while allowing transient response computation.

Nomenclature

\begin{itemize}
  \item $A,B,C,D =$ State space matrices
  \item $BZ =$ Block size of a subsystem
  \item $C =$ The set of all complex numbers
  \item $DSF =$ Downsampling factor
  \item $EM =$ Ending mode
  \item $I =$ Identity matrix
  \item $R =$ The set of all real numbers
  \item $SM =$ Starting mode
  \item $SS =$ Subsystem
  \item $TM =$ Total number of modes
  \item $T =$ CPU time
  \item $T =$ Sampling time period
  \item $Z =$ The set of all integers
  \item $m =$ Number of input channels
  \item $n =$ Number of samples
  \item $n_s =$ Number of state variables
  \item $p =$ Number of output channels
  \item $u =$ Control input
  \item $v =$ Discrete-time signal
  \item $w =$ Disturbance input
  \item $x =$ State vector
  \item $y =$ Measurement output
  \item $z =$ Performance output
  \item $\sigma =$ Root-mean-square
  \item $\theta =$ Attitude angles
  \item ./ =$ Matrix entry-wise division
  \item $d =$ Discrete-time / Disturbance
  \item $i =$ i-th subsystem
  \item $k =$ Controller
  \item $o =$ Optical controller
  \item $p =$ Plant
  \item $() =$ Lift operator
\end{itemize}

Subscripts

\begin{itemize}
  \item $d =$ Discrete-time / Disturbance
  \item $i =$ i-th subsystem
  \item $k =$ Controller
  \item $o =$ Optical controller
  \item $p =$ Plant
\end{itemize}

Superscripts

\begin{itemize}
  \item $-1 =$ Matrix inversion
  \item $T =$ Matrix transpose
\end{itemize}

1 Introduction

Simulation is an essential tool for the design of complex engineering systems, as the models of these systems become more and more complex. Fig. 1 shows some example systems and their complexity in terms of model dynamic bandwidth (normalized) and size. We have previously discussed the need for efficiency in large system time domain simulation and rationalized the distinction between large systems ($n_s \geq 500$), medium systems ($100 < n_s < 500$) and small systems ($n_s \leq 100$).
Fig. 1 Examples of size and dynamic bandwidth for LTI models of various space systems. The x-axis is the ratio between the highest and lowest natural frequencies in the model. The y-axis is the number of state variables of the model. The highlighted region motivates the presented simulation scheme.

Model dimension (or size/order) is a delicate issue in systems design because of the tradeoff between model fidelity and the number of designs explored within a limited time budget, $T_{tot} = N \times T_{cpu}$, where $N$ is the number of simulations or designs explored and $T_{cpu}$ is the runtime of a single simulation. The dilemma is shown in Fig. 2. It is clear from Fig. 2 that the more efficient a simulation is for a given level of model fidelity, the better a position system designers find themselves in. In addition, efficient simulation can facilitate the performance evaluation step of multidisciplinary design optimization (MDO), whose prospect and significance have been reported by Giesing et al.,

Sobieski et al.,

and Anderson among others. In Gutierrez, de Weck et al.,

and de Weck, three methods for performance evaluation of dynamic systems are summarized: 1) Time domain simulation, 2) Frequency domain analysis and 3) Lyapunov analysis. In these papers, improvements for frequency domain and Lyapunov methods were proposed with the derivation of new algorithms such as newlyap.m and modified internally balanced model reduction (see also Mallory with apriori error bounds, but the problem of time domain simulation remained unsolved. While frequency domain analysis and Lyapunov analysis can provide critical performance metrics such as steady state root-mean-square (RMS) values of performances, they cannot provide information on the transient response of the systems, which is sometimes required (e.g. in designing control systems). Also, while model reduction can result in smaller systems with acceptable accuracy, it suffers from the fact that arbitrary decisions must be made as to the amount of reduction and the balancing process itself requires substantial resources in computer time and memory, even though there are exciting improvements in this area, see for example, Willcox et al. and Beran. The problem just mentioned can be relieved without parallelization, if not entirely solved, by the capability of efficient time domain simulation of large order systems, which is the topic of this paper.

The proposed simulation algorithm, newlsim.m, first decouples the original dynamical system by implementing a block diagonalization of the A matrix, it then forms fictitious subsystems with lower, and thus manageable dimensions and narrower bandwidths. After that, the subsystems are discretized so that efficient computation of state transition can be realized. Finally the responses of the subsystems are interpolated and superposed to form the response of the original large-order system.

The organization of the paper is as follows: In Section 2 the technical time domain simulation problem will be discussed. Then in Section 3 the flow chart and some important details of the operations are presented. After describing the algorithm, Section 4 studies the simulation problems and solutions with various kinds of control loops. In Section 5 simulation results applied to a high fidelity Space Interferometry Mission (SIM) model, version 2.2, are given and some of these results are compared with those obtained by the standard MATLAB LTI systems simulator, lsim.m, as well as other techniques. A summary and conclusions are discussed in Section 6.

1For a diagonalizable matrix, the transformed matrix can be strictly diagonal but the result is usually complex, which is not very useful in practice. However, for a real diagonalizable matrix, the complex diagonal matrix can be converted into a real block diagonal one. Also, whenever the word “diagonal” is used in this paper, it means real block diagonal, unless noted otherwise.


2 Time Domain Simulation Problem

The time domain simulation problem of an LTI system can be defined as follows: Given a system in (1)

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

where \( x(t) \) is the state, \( u(t) \) is the input, \( y(t) \) is the output and the matrices are of appropriate dimensions. The solution to this problem is the unique \( y(t) \), given an external input \( u(t) \) and initial conditions \( x(t_0) \). There are at least two ways to solve the problem: 1) Standard ordinary differential equation (ODE) solvers like Runge Kutta, and 2) The state transition formula in (2)

\[
x(t) = e^{A(t-t_0)}x(t_0) + \int_{t_0}^{t} e^{A(t-\tau)}Bu(\tau)d\tau,
\]

where \( t_0 \) denotes the initial time instant, \( \tau \) is a dummy variable, \( e^{A(t-\tau)} \) is the matrix exponential of the matrix \( A(t-\tau) \) and \( u(t) \) is the external input to the system, \( x_H \) and \( x_P \) denote the homogenous and particular solutions, respectively. The state transition method just mentioned is equivalent to transforming the original problem (1) into its discretized version in (3), provided that a further assumption is made on \( u(t) \) (e.g. zero order hold (ZOH)).

\[
\begin{align*}
x[n+1] &= Ax[n] + Bu[n] \\
y[n] &= Cx[n] + Du[n],
\end{align*}
\]

where,

\[
\begin{align*}
A_d &= e^{AT}, \\
B_d &= \int_{t}^{T} e^{A(KT+T-\tau)}Bd\tau = \int_{0}^{T} e^{AT}Bd\tau,
\end{align*}
\]

and \( T \) is the sampling period and \( n \in \mathbb{Z} \).

The state transition method serves the current problem better in that it requires less computation and it maps left half s-plane poles to inside the unit circle in the z-plane, no matter how large the discretization time step, \( T \), is. Nevertheless, the benefits of the state transition method do not come for free in that the following problems must be addressed: The first problem is the computational expense of \( e^{AT} \) in (3). The cost of this operation is \( O(n_s^3) \) with \( n_s \) being the number of state variables.\(^{10} \) The second problem is memory saturation, if the computation requires the whole history of states before the history of the output can be computed (see \texttt{lsim.m} for such an algorithm\(^{11} \)), in that case the simulation might halt because of memory saturation. The solutions to these problems will be given in Section 3.

3 Fast Time Domain Simulation Algorithm

In this section, the steps and implementation issues of the presented algorithm will be addressed. Before the details are discussed, an overview of the algorithm is given by the flowchart in Fig. 3. In this flow chart,

...
divided into smaller subproblems. Also, this gives rise to the potential for parallel computation.

- **Applying multiple sampling rates.** The reason for this implementation is twofold. First, simulation can be facilitated by the application of multiple sampling rates. Secondly and more importantly, this can pave the way to solving multiple time scale dynamics problems, see Reich\textsuperscript{12} for such a problem in molecular dynamics.

- **Exploiting the sparsity resulting from the diagonal structure of the A matrix.** The number of nonzero entries of an ordinary dense matrix \( A \in \mathbb{R}^{n \times n} \) is \( n^2 \), but the number of nonzero entries in the diagonalized matrix is less then or equal to \( 2 \times n \). This sparsity is important in the matrix-vector product computation to be discussed later.

### Subsystem Planning

The subsystem planning subroutine is the key decision element in the algorithm. The objective of this function is to form fictitious subsystems and to assign them appropriate sampling rates. The assumption in this subsection is that the plant is already block diagonal and that the modes are sorted (i.e. after the preceding diagonalization step). There are two considerations for the subsystem planning:

The first issue is the size (number of state variables) of each subsystem. In terms of floating point operations (FLOPS), the size of the subsystems is not very important because the number of FLOPS for simulating discrete-time systems (after diagonalization) depends linearly on the number of state variables. According to Son\textsuperscript{10} the approximate cost is\textsuperscript{2},

\[
\text{FLOPS} = 2 \times (2 + m + p) \times n_s \times n \tag{4}
\]

where \( m \) is the number of input channels, \( p \) is the number of output channels, \( n_s \) is the number of state variables and \( n \) is the number of samples to be processed. This is in contrast to non-square continuous time system simulation, where the cost is proportional to the number of states to the third power.\textsuperscript{1} Nevertheless, simulating subsystems that are too large and too small is not efficient because of memory saturation and size independent overhead, respectively.

The other issue is the sampling rate for each simulation. The minimum sampling rate is given by Nyquist’s sampling theorem (e.g. see Oppenheim\textsuperscript{13}) but it is usually insufficient for computer simulations. As a rule of thumb, the sampling rate should be four to ten times the system bandwidth. The assumption in the preceding diagonalization step). There are two considerations for the subsystem planning:

The first issue is the size (number of state variables) of each subsystem. In terms of floating point operations (FLOPS), the size of the subsystems is not very important because the number of FLOPS for simulating discrete-time systems (after diagonalization) depends linearly on the number of state variables. According to Son\textsuperscript{10} the approximate cost is\textsuperscript{2},

\[
\text{FLOPS} = 2 \times (2 + m + p) \times n_s \times n \tag{4}
\]

where \( m \) is the number of input channels, \( p \) is the number of output channels, \( n_s \) is the number of state variables and \( n \) is the number of samples to be processed. This is in contrast to non-square continuous time system simulation, where the cost is proportional to the number of states to the third power.\textsuperscript{1} Nevertheless, simulating subsystems that are too large and too small is not efficient because of memory saturation and size independent overhead, respectively.

The other issue is the sampling rate for each simulation. The minimum sampling rate is given by Nyquist’s sampling theorem (e.g. see Oppenheim\textsuperscript{13}) but it is usually insufficient for computer simulations. As a rule of thumb, the sampling rate should be four to ten times the system bandwidth. The issue of appropriate sampling rate selection has been extensively discussed by Franklin\textsuperscript{14} and Aström\textsuperscript{15} among others.

Taking into account the aforementioned issues, the corresponding subsystem planning subroutine has the following features:

- It automatically chooses an appropriate subsystem block size by considering a lower bound, upper bound and the effect of the number of input and output channels.
- It automatically suggests downsampling based on an estimate of the ratio between the high frequency components and low frequency components of the output.

A flow chart of this subroutine is given in Fig. 4.

![Flow chart of the subsystem planning subroutine](image)

**Fig. 4 Flow chart of the subsystem planning subroutine.** DSF denotes downsampling factor (the ratio between the original sampling rate and the reduced sampling rate). SM is the starting mode of the subsystem. EM is the ending mode of the subsystem. TM is the total number of modes of the original system. This serves as a termination criterion. BZ is the currently chosen block size of each subsystem.

In this flowchart, the \( f(DSF) \) is implemented as a bisection process that determines the ending mode (EM) of each subsystem. As an example, the subsystem planning of the SIM v2.2 flexible mode model was considered with the planning result and downsampling factors given in Fig. 5.

### Discretization

As can be seen in (3), the bottleneck of discretization is the matrix exponential. Fortunately the diagonalization described earlier relieves the problem. By exploiting the sparsity of the block diagonal \( A \) matrix structure, a \( O(n_s) \) matrix exponential algorithm can be realized. This is obviously seen if the matrix exponential \( A_d \) in (3) is expressed as follows (see Chen,\textsuperscript{16} for example):

\[
A_d = e^{AT} = I + AT + \frac{A^2T^2}{2} + \frac{A^3T^3}{3!} + \cdots \tag{5}
\]
By noticing the fact that 
\[ A^n T^n = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_n \end{bmatrix} = \begin{bmatrix} A_1^n & 0 & \cdots & 0 \\ 0 & A_2^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_n^n \end{bmatrix} T^n, \]

where \( A_i \in \mathbb{R}^{2 \times 2} \forall i \in \{1, 2, \ldots\} \) and applying (6) to (5), the following equality holds:

\[ \exp\left( \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_n \end{bmatrix} \right) = \begin{bmatrix} e^{A_1 T} & 0 & \cdots & 0 \\ 0 & e^{A_2 T} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{A_n T} \end{bmatrix}. \]

Equation (7) is the basis of the proposed \( O(n_2) \) discretization scheme. As a result, a generic procedure for the fast discretization can be given as:

\begin{verbatim}
for i = 1 to total number of subsystems
    index = location of subsystem;
    DT_system(index) = c2d(subsystem);
end
\end{verbatim}

Here the size of each subsystem is not fixed and a theoretical optimal block size can be found. Nevertheless, the optimality is not an important issue since the difference between optimal and suboptimal strategies is not obvious here3.

A simple example as follows should reveal the fact1:

\begin{verbatim}
% create a random diagonal matrix
>> A = diag(abs(randn(1000,1)));
% dense matrix exponential
ans = 3.7303e-020
% computation time
elapsed_time = 27.1090
% diagonal matrix exponential
ans = 0.0310
% computation time
elapsed_time = 0.0310
% error checking
>> norm(Aexp - Aexp1)/length(Aexp)^2
ans = 3.7303e-020
\end{verbatim}

That is, the computation of the diagonal matrix exponential is 874 times faster than that of the dense matrix, and the difference between the results is negligible. Although in the example just shown, the \( A \) matrix is strictly diagonal, it can be concluded that a significant amount of computation is saved even in the case of a block diagonal \( A \) matrix, since the comparison is between \( O(n^3) \) and \( O(n) \) operations. Nevertheless, the advantage of fast discretization comes with the expense of the diagonalization step itself, which is also \( O(n^3) \). Fortunately, the computation of eigenvalues is usually more efficient than that of the matrix exponential, as Table 1 shows.

<table>
<thead>
<tr>
<th>Size (n)</th>
<th>100</th>
<th>200</th>
<th>400</th>
<th>600</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>eig</td>
<td>0.0310</td>
<td>0.2340</td>
<td>1.9850</td>
<td>6.4530</td>
<td>29.4530</td>
</tr>
<tr>
<td>expm</td>
<td>0.0310</td>
<td>0.3280</td>
<td>2.6250</td>
<td>9.0620</td>
<td>40.9530</td>
</tr>
</tbody>
</table>

Downsampling

Downsampling the input is required whenever the assigned sampling rate of a particular simulation is lower than that of the input (this is true if the multiple sampling rates strategy is employed). The straightforward way to achieve the goal is to low-pass filter the signal and then downsample it (as Oppenheim, for example13). That is, 

\[ x_d[n] = x[Mn], \]

where \( x[n] \) has been low-pass filtered and \( M \) is the downsampling factor.

In addition to the direct method just discussed, there is another way to downsample. Recall the state transition formula (or the state equations for the discretized system),

\[ \begin{align*}
    x[n + 1] &= A_d x[n] + B_d u[n] \\
    y[n] &= C x[n] + D u[n].
\end{align*} \]

The first equation in (8) is certainly satisfied at the \( n + 1 \) instant,

\[ x[n + 2] = A_d x[n + 1] + B_d u[n + 1]. \]
If (8) is substituted into (9), then the following results
\[
x[n + 2] = A_d \{ A_d x[n] + B_d u[n] \} + B_d u[n + 1] \\
= A_d^2 x[n] + A_d B_d u[n] + B_d u[n + 1].
\]
(10)

Equation (10) can be generalized for N steps,
\[
x[n + N] = A_d^N x[n] + A_d^{N-1} B_d u[n] + \cdots + B_d u[n + N - 1]
\]
(11)

If the following new matrices are defined
\[
u[n] = \begin{bmatrix} u[n] & \cdots & u[n + N - 1] \end{bmatrix}^T \\
A_d = A_d^N \\
B_d = \begin{bmatrix} A_d^{N-1} & B_d & \cdots \\ & & B_d \end{bmatrix} \\
C = C \\
D = \begin{bmatrix} D & 0 & \cdots & 0 \end{bmatrix},
\]
then the N-step propagation version of (8) is
\[
x[n + N] = A_d x[n] + B_d u[n] \\
y[n] = C x[n] + D u[n].
\]
(12)

By employing a long time step state transition, the calculation of the unwanted intermediate states and outputs can be avoided. This downsampling scheme essentially downsamples the output instead of the input and the accuracy is much better than the first direct approach. However, the efficiency gain achieved by the second method is less than that by the first method. Compare the FLOPS for the direct downsampling (13) and those of the second method (14):
\[
\text{FLOPS} = 2 \times \frac{(2 + m + p) \times n_s \times n}{DSF},
\]
(13)
\[
\text{FLOPS} = 2 \times \frac{(2 + DSF \times m + p) \times n_s \times n}{DSF}.
\]
(14)

Here \( DSF \geq 1 \) is the downsampling factor. The meanings of other parameters in (13) and (14) are referred back to (4). In conclusion, the second method is more conservative. The current version of newlsim.m adopts the second method, also referred to as lifting, if downsampling is to be realized.

Simulation, Interpolation and Superposition

With the subsystems formed and discretized, the burden on the simulator (the actual code that computes the states and outputs) is lighter compared to the original ODE problem and it now becomes a much simpler problem of matrix-vector multiplication (cf. (3)). Taking into account the sparsity of the \( A \) matrix, the matrix-vector multiplication (state transition) can be realized in \( O(n_s n) \) (\( n_s \) is the number of state variables and \( n \) is the number of samples to be simulated) FLOPS. The required features of the simulator are summarized as follows:

- The simulator must be memory conscious. It cannot request any amount of memory (in bytes) proportional to \( n_s n \) or more.
- The simulator must be able to recognize the zero pattern of the \( A \) matrix, otherwise, the advantage of the sparsity will be lost and the computation effort estimate in (4) will not be achieved.

Taking into account the above considerations, MATLAB SIMULINK’s \( \text{dsf} \) discrete-time system solver “discrete state-space” is chosen, instead of the conventional choice of \( \text{ltitr} \).

Interpolation is needed whenever the output is downsampled. Because the downsampling instants are evenly spaced, the downsampling can be viewed as the resampling of discrete-time signals and efficient interpolation methods in signal processing (i.e. inserting zeros and then low-pass filtering) can be employed. In fact, interp.m applies the MATLAB command interp.m, which does exactly this.

With the responses of the subsystems computed, it is finally possible to form the response of the original system due to the original input. This is allowable because of the linearity property of LTI systems.

The commands mentioned above are very MATLAB specific because the current version of newlsim.m is implemented in MATLAB. Nevertheless, the idea to take advantage of problem specific insights and structure extends naturally to more general platforms.

4 Closed Loop Systems Issues

In this section, issues concerning the implementation of the newlsim.m algorithm are addressed. The first problem is due to closing a feedback loop (e.g. attitude control systems (ACS) for a satellite). The second problem is due to cascade connections between the plant and some other filters (e.g. noise shaping filter and/or optical controller). These two problems will be studied in two separate subsections.

Simulation with Feedback Loop

The block diagram of the problem is given in Figure 6.

```
\begin{center}
\begin{tikzpicture}[scale=1,auto,>=latex]
\node [input, name=w] at (0,0) {$w$};
\node [state, name=x] at (2,0) {$A_k B_k \quad C_k \quad D_k$};
\node [input, name=y] at (2,-1.5) {$y$};
\node [input, name=u] at (2,-3) {$u$};
\node [input, name=p] at (2,-4.5) {};\node [input, name=p] at (2,-4.5) {past};
\node [state, name=controller] at (2,-6) {$A_k \quad B_k \quad C_k \quad D_k$};
\node [state, name=plant] at (2,-7.5) {$A_k \quad B_k \quad C_k \quad D_k$};
\draw [->] (w) -- (x);
\draw [->] (x) -- (y);
\draw [->] (u) -- (x);
\draw [->] (y) -- (controller);
\draw [->] (p) -- (controller);
\end{tikzpicture}
\end{center}
```

Fig. 6 Block diagram of the plant with a feedback controller.

Suppose the open loop system is in modal form (i.e. the \( A \) matrix is block diagonal) and the state-space
The state space form is:
\[
\begin{bmatrix}
A & B_w & B_u \\
C_z & D_{zw} & D_{zu} \\
C_y & D_{yw} & 0
\end{bmatrix} .
\] (15)

The feedback controller has the following state-space realization:
\[
\begin{bmatrix}
A_k & B_k \\
C_k & D_k
\end{bmatrix} ,
\] (16)

where subscripts \(w\) and \(u\) denote quantities related to disturbance input and control input, respectively. Subscripts \(z\) and \(y\) are related to performance output and measurement output, and subscript \(k\) denotes quantities related to the controller. The closed loop system has the state space form
\[
\begin{bmatrix}
A + B_u D_k C_y & B_u C_k & B_w + B_u D_k D_{yw} \\
B_k C_y & A_k & B_k D_{yw} \\
C_z + D_{zw} D_k C_y & D_{zu} C_k & D_{zw} + D_{zu} D_k D_{yw}
\end{bmatrix} .
\] (17)

The off-diagonal terms, \(B_u C_k\) and \(B_k C_y\) in the “\(A\)” matrix of (17) are not expected to be zero, otherwise there will be no control effect at all. Now the problem is: Even if the open loop system is in modal form (i.e. the \(A\) matrix is block diagonal), the closed loop system will not be so because of the dynamics coupling (off diagonal terms in “\(A\)” in (17)). Another problem arises if the feedback controller is a discrete-time system\(^6\) and this causes the closed loop system to be hybrid\(^7\). For the problem of dynamics coupling, two solutions are proposed:

**Rediagonalization by eigenvalue decomposition.** An eigenvalue decomposition is applied to the system in (17). This is the most straightforward way but the computation can be expensive if the systems considered are large-order, because of the eigenvalue problem involved.

**Forced decoupling.** This is a heuristic method in that some of the entries of \(C_y\) in (17) are set to zero. In other words, some of the measurements are regarded as insensitive to some of the state variables. Suppose for simplicity that all \(D\) matrices are zero and the state variables are reordered in such a way that the following equalities hold (\(A\) is assumed to be block diagonal and \(C_y\) is partitioned into two blocks):

\[
A \begin{bmatrix}
A_1 & 0 \\
0 & A_2
\end{bmatrix} .
\]

\(^3\)The \(D_{yw}\) is missing here to avoid an algebraic loop, i.e. the coexistence of feedthroughs in the plant and the controller.

\(^6\)In practical implementations, controllers are usually digital, which implies that the signals, as well time instants, are discrete. The discrete-time assumption here is merely for the ease of analysis.

\(^7\)Here the term “hybrid” is used in the very specific sense that the system contains both continuous-time and discrete-time states. There is no discrete state involved.

Then the state-space representation of the closed loop system (17) is as follows:
\[
\begin{bmatrix}
A_1 & 0 & B_{1w} \\
0 & A_2 & B_{2w} \\
0 & B_k C_{2y} & A_k \\
C_{1z} & C_{2z} & 0
\end{bmatrix} .
\] (18)

It can be verified that the system in (18) can be decomposed into two subsystems. The first subsystem includes controller dynamics and is subject to disturbance input only,
\[
\begin{bmatrix}
A_2 & B_{2w} C_k \\
B_k C_{2y} & A_k \\
C_{2z} & 0
\end{bmatrix} (19)
\]

and the second subsystem evolves in time with disturbance input and control input that is determined by solving the first subsystem (as the output signal of the controller)
\[
\begin{bmatrix}
A_1 & B_{1w} \\
0 & 0
\end{bmatrix} .
\] (20)

It can be seen that if the dimension of \(A_2\) in (19) is much smaller than that of \(A_1\) in (20) and if \(A_1\) is block diagonal, then the bottleneck of diagonalization can be avoided. The justification of this method hinges upon the ability to find the state variables that are insensitive to sensor measurements and the relative significance of the contributions of the ignored measurements to the total measurements. The determination of the “important” state variables can be quite case specific. For example, in the study of a space structure with an attitude control system (ACS), if the measurements are attitude angles, then it is natural that the rigid body modes are far more important than other flexible modes. In order to quantify the error induced by the forced decoupling method, it is possible to compute the ratio
\[
E = \frac{\sigma_1}{\sigma_2} ,
\] (21)

where \(\sigma_1\) and \(\sigma_2\) are the open (feedback control) loop RMS values of the contributions of the unimportant
and important dynamics to the measurement. If $E$ is smaller than some tolerance, then the forced decoupling heuristic is justified, otherwise rediagonalization by eigenvalue decomposition must be applied. The computation of the ratio $E$ in (21) can be very efficient if the open loop system is already in modal form. For example, the Lyapunov analysis with newlyap.m in de Weck et al.\(^1\) can be applied here. As an example, consider the 2184 state variable SIM model v2.2 with three unstable (or marginally stable) rotational rigid body modes and assume the attitude angles (proportional to rigid body mode angles, together with some additional contributions from other flexible modes) are measured directly. The numeric values of $E$ in (21) in this example are summarized in Table 2.

**Table 2** $E$: The ratio (%) between the RMS attitude angle of the flexible mode subsystem over the rigid body mode subsystem.

<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1021</td>
<td>0.2201</td>
<td>0.1950</td>
</tr>
</tbody>
</table>

The results in Table 2 shows that $E$ is quite small if the forced decoupling heuristic is used. To verify the prediction, the actual results from the two methods are computed: The RMS values of the performance outputs by eigenvalue decomposition method and the forced decoupling heuristic are 1.8275 × 10^{-5} and 1.8276 × 10^{-5}, respectively. Their difference is 1.8135 × 10^{-5}, which amounts to about 0.0099 % of the performance given by the eigenvalue decomposition method (chosen as a reference here). In conclusion, the forced decoupling heuristic is not exact but can be fairly accurate if properly applied.

For the problem of “hybrid” closed loop systems, there are two options suggested:

**Continuous-time approximation.** If the sampling rate of the control is high enough, then the digital controller can actually be approximated by a continuous-time system using techniques like zero order hold (ZOH) or bilinear (Tustin) transform. The accuracy of this approximation can be found in any common digital (or computer) control text.

**Lifting.** This is a useful approach to deal with sampled-data control systems analysis problems. For example, see Chen\(^19\) and Yamamoto.\(^20\) Suppose a discrete-time signal $v[n]$ is defined as

$$v = \{v[0], v[1], v[2], \ldots\}.$$  

The lifted version of the signal $v$ can be expressed as

$$v = \left[ \begin{array}{c} v[0] \\ v[1] \\ \vdots \\ v[n-1] \\ v[n] \\ v[n+1] \\ \vdots \\ v[2n-1] \end{array} \right].$$

where $n \in \mathbb{Z}$. If an LTI system is represented as

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

then its lifted version is defined such that the original inputs and outputs signals are lifted. That is,

$$\begin{bmatrix} A^n & A^{n-1}B & A^{n-2}B & \cdots & B \\ C & D & 0 & \cdots & 0 \\ CA & CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{n-1} & CA^{n-2}B & CA^{n-3}B & \cdots & D \end{bmatrix}$$  \hspace{1cm} (22)

The lifting procedure in (22) basically reduces sampling rate at the expense of an increase in input and output dimensions. Equivalently this procedure can be viewed as one of the applications of the state augmentation technique. The main application of this method in the algorithm is to convert a multiple sampling rates\(^8\) system into a single rate (slow rate) LTI system without losing the effect of fast dynamics. The drawback of this method is the high resulting dimensionality. Nevertheless, this method can work well in conjunction with the forced decoupling method discussed previously if the subsystem coupled with the controller has low dimension.

**Simulation with Feedforward Controller**

The block diagram is depicted in Figure 7. Even though the open loop plant is in modal form (block diagonal $A$ matrix), the cascade connection of the plant and other dynamic systems (e.g. noise shaping filter or feedforward controller) does not in general have a block diagonal $A$ matrix. Suppose the plant dynamics is

$$\begin{bmatrix} A_p & B_p \\ C_p & 0 \end{bmatrix},$$  \hspace{1cm} (23)

and the controller has the state-space representation as

$$\begin{bmatrix} A_k & B_k \\ C_k & 0 \end{bmatrix},$$  \hspace{1cm} (24)

\(\text{\footnote{Sometimes the sampling rate of the plant is much higher than that of the controller, in order to represent the plant dynamics accurately, see Chen\(^19\) for more detail. Note also that a multiple sampling rate system is time-variant.}}\)
then the cascade connection of these two systems will have the state-space representation as follows:
\[
\begin{bmatrix}
A_p & B_pC_k & 0 \\
0 & A_k & B_k \\
C_p & 0 & 0
\end{bmatrix}.
\] (25)

It is clear that even if \(A_p\) is block diagonal, the \(A\) matrix of the closed loop system will not be so because of \(B_pC_k\). Therefore, a rediagonalization is necessary. Nevertheless, it is not necessary to call the eigenvalue solver (e.g. \texttt{eig} in MATLAB) to redo the diagonalization if the plant and the controller do not have the same eigenvalues. That is,
\[
\lambda(A_p) \cap \lambda(A_k) = \emptyset,
\]
where \(\lambda(\cdot)\) denotes the set of all eigenvalues of a matrix. This is true because of the following lemma (from Golub and Van Loan21):

**Lemma 4.1** Let \(T \in \mathbb{C}^{n \times n}\) be partitioned as follows:
\[
T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix},
\] (26)

where \(T_{11} \in \mathbb{C}^{p \times p}\) and \(T_{22} \in \mathbb{C}^{q \times q}\). Define the linear transformation \(\phi : \mathbb{C}^{p \times q} \rightarrow \mathbb{C}^{p \times q}\) by
\[
\phi(X) = T_{11}X - XT_{22},
\] (27)
where \(X \in \mathbb{C}^{p \times q}\). Then \(\phi\) is nonsingular if and only if \(\lambda(T_{11}) \cap \lambda(T_{22}) = \emptyset\), where \(\lambda(T_{11})\) and \(\lambda(T_{22})\) are sets of eigenvalues of \(T_{11}\) and \(T_{22}\) respectively. If \(\phi\) is nonsingular and \(Y\) is defined by
\[
Y = \begin{bmatrix} I_p & Z \\ 0 & I_q \end{bmatrix},
\]
then \(Y^{-1}TY = \text{diag}(T_{11}, T_{22})\).

**Proof:** See Appendix.

In the current context, \(A_p\) and \(A_k\) can be thought of as \(T_{11}\) and \(T_{22}\) in the lemma. It is assumed that the diagonalization of \(A_k\) (or \(T_{22}\)) is possible and easy to find (e.g. controller A matrix). The implication of this lemma is that the eigenvalues and eigenvectors of the closed loop A matrix can be found without using general purpose eigenvalue solvers (e.g. \texttt{eig} in MATLAB), and thus the rediagonalization can be computed efficiently. The reason is as follows: Suppose \(T_{11} \in \mathbb{R}^{p \times p}\) and \(T_{22} \in \mathbb{R}^{q \times q}\) are block diagonal and diagonalizable, so there exist invertible matrices (eigenvector matrices) \(V_{11}\) and \(V_{22}\) such that \(V_{11}^{-1}T_{11}V_{11} = D_{11}\) and \(V_{22}^{-1}T_{22}V_{22} = D_{22}\), where \(D_{11} \in \mathbb{C}^{p \times p}\) and \(D_{22} \in \mathbb{C}^{q \times q}\) are both strictly diagonal. The matrix \(T\) in (26) can be expressed as
\[
\begin{bmatrix}
V_{11}^{-1} & 0 \\
0 & V_{22}^{-1}
\end{bmatrix} \begin{bmatrix} D_{11} & D_{12} \\ 0 & D_{22} \end{bmatrix} \begin{bmatrix} V_{11} & 0 \\ 0 & V_{22} \end{bmatrix},
\] (28)
where \(D_{12} = V_{11}T_{12}V_{22}^{-1}\). The middle matrix in (28) can be transformed into block diagonal form by applying lemma 4.1. That is,
\[
\begin{bmatrix} D_{11} & D_{12} \\ 0 & D_{22} \end{bmatrix} = \begin{bmatrix} I & X' \\ 0 & I \end{bmatrix} \begin{bmatrix} D_{11} & 0 \\ 0 & D_{22} \end{bmatrix} \begin{bmatrix} I & -X' \\ 0 & I \end{bmatrix},
\] (29)

where \(X'\) can be solved by the following equation (essentially (27))
\[
-D_{12} = D_{11}X' - X'D_{22}.
\]

It can be seen that
\[
X' = -D_{12} \cdot / (D_{11}U - UD_{22}),
\] (30)

where \(U\) denotes a matrix whose entries are all one and \(\cdot /\) denotes the operation of elementwise division. Combining (29) and (30), it can be seen that the eigenvector matrix of \(T\) in (26) is
\[
\begin{bmatrix} I & -X' \\ 0 & I \end{bmatrix} \begin{bmatrix} V_{11} & 0 \\ 0 & V_{22} \end{bmatrix} = \begin{bmatrix} V_{11} & -X'V_{22} \\ 0 & V_{22} \end{bmatrix}.
\] (31)

With the eigenvalues and eigenvectors of \(T\) in (26) known, it is straightforward to compute the rediagonalization required in the beginning of this subsection.

**5 Simulation Results**

In this section, some application examples are given to show the potential value of the \texttt{newlsim.m} algorithm. The first example computes the performance RMS values of randomly generated stable SISO systems of different dimensions, \(n_s\), driven by randomly generated input signals. As in de Weck et al., three methods are compared: Time domain method, frequency domain method and Lyapunov method. For the time domain method, the input signal is \(10^5\) samples long and for the frequency domain method, \(10^5\) frequency points (single sided) are computed. The results are summarized in Table 3.

In this table, \(n_s\) is the number of state variables. \(T_{cpu}\) is the CPU time (in seconds) of each computation and \(\sigma\) is the RMS value of each performance output (the actual unit is not of interest here). Methods: \texttt{freq} denotes the frequency domain method, \texttt{newlyap} denotes the fast Lyapunov method, \texttt{lsim} is the standard time domain simulation method provided by MATLAB. Finally, \texttt{newlsim} is the proposed time domain simulation method. Note that \texttt{freq} and \texttt{newlyap} are the fast implementations of frequency domain method and Lyapunov method respectively, see de Weck et al.\ Note also that the time for \texttt{newlsim} includes the time to diagonalize. Since new data is generated in each case with a different \(n_s\), the RMS values of different \(n_s\) cases differ accordingly and should be not compared.

The main point to illustrate here is that the performance RMS values computed by the three different
methods are quite close to each other. As shown in Table 3, time domain methods are generally not as efficient as other methods for small systems ($n_s < 100$). However, the situation changes when systems get larger ($n_s > 50$). The reason for this trend is that size-independent overhead of the time domain method is more significant in small system cases. It should also be noted that the computation time of the frequency domain and time domain methods varies with the number of samples evaluated. Nevertheless, it is this computation that provides the additional information that the Lyapunov method does not provide (i.e. time history and power spectral density). The reason why the results from \texttt{lsim} are unavailable (N/A) is that the computer ran out of memory, which means that \texttt{lsim} is not very suited for large-order systems simulation. The above example shows that \texttt{newlsim.m} can achieve efficiency similar to the fastest implementations of other methods (frequency domain and Lyapunov methods) with acceptable accuracy when computing RMS values.

Computing output RMS values does not fully exemplify the advantage of \texttt{newlsim.m}. A time domain simulation scheme should be compared with another time domain simulation scheme. Therefore the MATLAB simulator, \texttt{lsim.m}, is chosen as a reference in the following example. In the example, a number of randomly generated systems with different sizes are simulated in the time domain with \texttt{lsim.m} and \texttt{newlsim.m}. The computation time of each simulation is given in Figure 8. The time responses of a sample system by \texttt{lsim.m} and \texttt{newlsim.m} are shown in Figure 9.

The example shows that \texttt{newlsim.m} is more efficient than \texttt{lsim.m}, while the error is insignificant.

The remaining examples are concerned with control tuning of the Space Interferometry Mission (SIM) model v2.2 (see Figure 10 for its finite element model) that is enabled by \texttt{newlsim.m}.

The SIM v2.2 model presents significant challenges.
to time domain simulation because of its high dimensionality (2184 state variables) and wide dynamic range \((\omega_{\text{min}}/\omega_{\text{max}} \approx 4700)\). For more information on SIM, see JPL.\textsuperscript{22} Due to the scientific purpose, stringent design requirements are imposed on the SIM model. For example, see Table 4 for some requirements from Miller et al.\textsuperscript{23} and Laskin.\textsuperscript{24}

**Table 4** SIM opto-mechanical performance requirements.

<table>
<thead>
<tr>
<th>Performance</th>
<th>Unit</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starlight OPD</td>
<td>nm</td>
<td>10 (RMS)</td>
</tr>
<tr>
<td>Internal Metrology OPD</td>
<td>nm</td>
<td>20 (RMS)</td>
</tr>
<tr>
<td>Starlight WFT</td>
<td>s</td>
<td>0.210 (RSS)</td>
</tr>
</tbody>
</table>

In order for the SIM system to work properly, two control systems are needed. One is the attitude control system (ACS) and the other is the optical control system. The overall system configuration is given in Fig. 11.

![Fig. 10 Finite element model of SIM v2.2.](image)

The ACS in Fig. 11 stabilizes the open loop unstable rigid body modes of the SIM model. It can be designed by classical methods like PID, lead-lag or modern control techniques such as LQG. The optical control here is modeled as a second order high-pass filter and the transfer function of one channel is

\[
K_o(s) = \frac{s^2}{s^2 + 2\zeta_oo_s + \omega_o^2},
\]

where \(\zeta_o\) is the damping ratio of the controller and is set to 0.707 and \(\omega_o\) is the corner frequency, which is treated as a variable design parameter.

The first example is a parameter study of optical controller corner frequency \(\omega_o\) [rad/s] (or \(f_o\) [Hz]). The system consists of the open loop SIM model (2184 state variables), an ACS designed by the standard LQG approach (e.g. Bélanger\textsuperscript{25}) and the optical controller as given in (32). The ACS loop is closed by a rediagonalization by eigenvalue decomposition and the optical control path is closed by the method prescribed in Section 4. There are six input channels (three forces and three torques), which are driven by the six channels of Magellan reaction wheel assembly disturbance data (see Elias\textsuperscript{26}). The outputs are starlight optical path difference (OPD), internal metrology (IM) and starlight wavefront tilt (WFT). In the simulation runs, different closed loop systems with different optical controller corner frequencies \(f_o\) are formed and the RMS values of the performance outputs are recorded. The result is shown in Figure 12. The result in Figure 12 is consistent with the intuition that higher optical control bandwidth leads to better system performance. Nevertheless, it can be a problem of cost, implementation and stability margins if \(f_o\) is chosen too high. A controller cutoff frequency above 10 [Hz] appears to satisfy the requirements.

The second example is the tuning of the attitude control system (ACS). The performance outputs here are the three attitude angles \((\theta_x, \theta_y, \text{ and } \theta_z)\). In this
tuning, two typical scenarios are shown. One is the cheap control case and the other is the expensive control case. These cases are determined by the weights on the state and control when the ACS controller is designed. The unit step transient responses of the three attitude angles are shown in Figure 13.

It can be seen from Figure 13 that while the settling time reduces with the increase of control effort, the overshoot remains large. The reason for this difficulty is the existence of non-minimum phase zeros of the open loop plant. In the simplest interpretation, the non-minimum phase zeros draw the closed loop root loci further to right half s-plane as control effort increases, and are thus reducing the stability margin. Therefore, if a structural design such that the non-minimum phase zeros are eliminated can be realized, then the transient responses of the model are expected to be much better. These transient analyses for large systems require fast time domain simulation algorithms such as `newlsim.m`.

6 Summary

Time domain simulation is an important technique for multidisciplinary design, analysis and optimization of dynamic systems. Unfortunately, as model fidelity and size get large one experiences excessive computation times and memory saturation problems. In this paper, the simulation scheme, `newlsim.m`, based on a block diagonalization pre-processing, is presented in response to this challenge. The targeted systems are large-order, diagonalizable continuous-time LTI systems. It has been found that diagonalization provides three benefits (dynamics decoupling, fast discretization and multiple sampling rates) that facilitate the simulation. In conjunction with the block diagonalization structure of the resultant A matrix, it has been shown that a sparse matrix recognizable state transition must be employed in order to achieve the O(n^2) state transition by taking advantage of the resultant sparsity. Problems with feedback and feedforward controllers are discussed and the corresponding solutions (e.g., forced decoupling and rediagonalization without using iterative eigenvalue solver) are proposed. It has been shown that `newlsim.m` can achieve similar efficiencies as those achieved by fast implementations of frequency domain and Lyapunov methods (e.g., `newlyap.m`), while retaining the advantage of transient response calculations. Finally, applications enabled by `newlsim.m` are given as optical and ACS control tuning of the 2200 state SIM spacecraft system to illustrate the potential value of the simulation scheme.

Recommendations for future work include extensions of the algorithm to time-varying and weakly nonlinear systems as well implementation of distributed computation of the subsystem responses on parallel computers.

Acknowledgement

This research was funded by the NASA Jet Propulsion Laboratory under grant No. JPL 91123 and was monitored by Dr. Ipek Basdogan. The authors thank Prof. Wallace Vander Velde of MIT for his helpful comments.

Appendix: Derivation of Lemma 4.1

Suppose \( \phi(X) = 0 \) for \( X \neq 0 \) and that

\[
U^H XV = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r \\ -r \end{bmatrix}
\]

is the SVD of \( X \) with \( \Sigma_r = \text{diag}(\sigma_i) \), \( r = \text{rank}(X) \). Substituting this into the equation \( T_{11}X = XT_{22} \) gives

\[
\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},
\]

where \( U^HT_{11}U = (A_{ij}) \) and \( V^HT_{22}V = (B_{ij}) \). By comparing blocks we see that \( A_{21} = 0, B_{12} = 0, \) and \( \lambda(A_{11}) = \lambda(B_{11}) \). Consequently,

\[
\emptyset \neq \lambda(A_{11}) = \lambda(B_{11}) \subseteq \lambda(T_{11}) \cap \lambda(T_{22}).
\]

On the other hand, if \( \lambda \in \lambda(T_{11}) \cap \lambda(T_{22}) \) then we have nonzero vectors \( x \) and \( y \) so \( T_{11}x = \lambda x \) and \( y^HT_{22} = \emptyset \).
A calculation shows that $\phi(xy^H) = 0$. Finally if $\phi$ is nonsingular then the matrix $Z$ above exists and

$$Y^{-1}TY = \begin{bmatrix} I & -Z \\ 0 & I \end{bmatrix} \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \begin{bmatrix} I & Z \\ 0 & I \end{bmatrix} = \begin{bmatrix} T_{11} & 0 \\ 0 & T_{22} \end{bmatrix}.$$ 

Q.E.D.

References


