SINGULAR PERTURBATIONS
AND TIME-SCALE METHODS
IN CONTROL THEORY:
SURVEY 1976-1982

V. R. SAKSENA
J. O'REILLY
P. V. KOKOTOVIC

APPROVED FOR PUBLIC RELEASE. DISTRIBUTION UNLIMITED.

REPORT R-979
UILU-ENG 82-2245

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Recent progress in the use of singular perturbation and two-time-scale methods of modeling and design for control systems is reviewed. Over 400 references are organized into major problem areas. Representative issues and results are discussed with a view to outlining research directions and indicating potential areas of application. The survey is aimed at engineers and applied mathematicians interested in model-order reduction, separation of time scales and allied simplified methods of control system analysis and design. The exposition does not assume prior knowledge of singular perturbation methods.
Recent progress in the use of singular perturbation and two-time-scale methods of modeling and design for control systems is reviewed. Over 400 references are organized into major problem areas. Representative issues and results are discussed with a view to outlining research directions and indicating potential areas of application. The survey is aimed at engineers and applied mathematicians interested in model-order reduction, separation of time scales and allied simplified methods of control system analysis and design. The exposition does not assume prior knowledge of singular perturbation methods.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1. TWO-TIME-SCALE SYSTEMS</td>
<td>4</td>
</tr>
<tr>
<td>1.1. Linear Systems</td>
<td>5</td>
</tr>
<tr>
<td>1.2. Nonlinear Systems</td>
<td>9</td>
</tr>
<tr>
<td>1.3. Discrete Time Systems</td>
<td>11</td>
</tr>
<tr>
<td>2. SYSTEM PROPERTIES</td>
<td>14</td>
</tr>
<tr>
<td>2.1. Stability</td>
<td>14</td>
</tr>
<tr>
<td>2.2. Controllability and Observability</td>
<td>16</td>
</tr>
<tr>
<td>2.3. Analysis in the Frequency Domain</td>
<td>17</td>
</tr>
<tr>
<td>3. LINEAR COMPOSITE CONTROL</td>
<td>20</td>
</tr>
<tr>
<td>3.1. Linear State Feedback</td>
<td>20</td>
</tr>
<tr>
<td>3.2. Output Feedback and Observers</td>
<td>22</td>
</tr>
<tr>
<td>4. NONLINEAR AND ADAPTIVE CONTROL</td>
<td>25</td>
</tr>
<tr>
<td>4.1. Nonlinear Composite Control</td>
<td>25</td>
</tr>
<tr>
<td>4.2. Trajectory Optimization</td>
<td>28</td>
</tr>
<tr>
<td>4.3. Adaptive Control</td>
<td>30</td>
</tr>
<tr>
<td>5. STOCHASTIC FILTERING AND CONTROL</td>
<td>33</td>
</tr>
<tr>
<td>5.1. LQG Problems</td>
<td>33</td>
</tr>
<tr>
<td>5.2. Nonlinear Stochastic Control</td>
<td>38</td>
</tr>
<tr>
<td>5.3. Wide-Band Input Noise Formulations</td>
<td>40</td>
</tr>
<tr>
<td>6. HIGH-GAIN FEEDBACK SYSTEMS</td>
<td>43</td>
</tr>
<tr>
<td>6.1. High-Gain and Time Scales</td>
<td>43</td>
</tr>
<tr>
<td>6.2. Composite State Feedback and Cheap Control</td>
<td>45</td>
</tr>
<tr>
<td>6.3. Variable Structure Systems</td>
<td>47</td>
</tr>
<tr>
<td>7. WEAK COUPLING AND TIME SCALES</td>
<td>49</td>
</tr>
<tr>
<td>7.1. Slow Coherency in Power Systems</td>
<td>49</td>
</tr>
<tr>
<td>7.2. Aggregation of Markov Chains</td>
<td>51</td>
</tr>
<tr>
<td>8. CONTROL OF LARGE SCALE SYSTEMS</td>
<td>53</td>
</tr>
<tr>
<td>8.1. Multimodeling</td>
<td>53</td>
</tr>
<tr>
<td>8.2. Singly Perturbed Differential Games</td>
<td>57</td>
</tr>
<tr>
<td>CONCLUDING REMARKS</td>
<td>60</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>61</td>
</tr>
<tr>
<td>------------</td>
<td>----</td>
</tr>
<tr>
<td>A) Books and Surveys</td>
<td>61</td>
</tr>
<tr>
<td>B) Asymptotic Methods</td>
<td>62</td>
</tr>
<tr>
<td>C) Time-Scale Properties of Linear Systems</td>
<td>64</td>
</tr>
<tr>
<td>D) Discrete Time Systems</td>
<td>65</td>
</tr>
<tr>
<td>E) Linear State Feedback</td>
<td>67</td>
</tr>
<tr>
<td>F) Output Feedback and Observers</td>
<td>70</td>
</tr>
<tr>
<td>G) Stability, Non-Linear and Adaptive Control</td>
<td>71</td>
</tr>
<tr>
<td>H) Trajectory Optimization</td>
<td>73</td>
</tr>
<tr>
<td>I) Stochastic Systems, Filtering and Control</td>
<td>75</td>
</tr>
<tr>
<td>J) High-Gain Feedback Systems</td>
<td>77</td>
</tr>
<tr>
<td>K) Time Scales in Networks and Power Systems</td>
<td>81</td>
</tr>
<tr>
<td>L) Aggregation of Markov Chains</td>
<td>82</td>
</tr>
<tr>
<td>M) Multimodeling and Stability of Large Scale Systems</td>
<td>83</td>
</tr>
<tr>
<td>N) Singularity Perturbed Differential Games</td>
<td>85</td>
</tr>
<tr>
<td>O) Singular Systems</td>
<td>86</td>
</tr>
<tr>
<td>Additional References</td>
<td>88</td>
</tr>
</tbody>
</table>
INTRODUCTION

Singularly perturbed systems and, more generally, two-time-scale systems, often occur naturally due to the presence of small "parasitic" parameters, typically small time constants, masses, etc., multiplying time derivatives or, in more disguised form, due to the presence of large feedback gains and weak coupling. The chief purpose of the singular perturbation approach to analysis and design is the alleviation of the high dimensionality and ill-conditioning resulting from the interaction of slow and fast dynamic modes. This two-time-scale approach is asymptotic, that is, exact in the limit as the ratio $\varepsilon$ of the speeds of the slow versus the fast dynamics tends to zero. When $\varepsilon$ is small, approximations are obtained from reduced-order models in separate time scales.

While singular perturbation theory, a traditional tool of fluid dynamics and nonlinear mechanics, embraces a wide variety of dynamic phenomena possessing slow and fast modes, its assimilation in control theory is recent and rapidly developing. The methods of singular perturbations for initial and boundary value problem approximations and stability were already largely established in the 1960s, when they first became a means for simplified computation of optimal trajectories. It was soon recognized that singular perturbations are present in most classical and modern control schemes based on reduced order models which disregard high frequency "parasitics." This recognition led to the development of two-time-scale methods for a variety of applications including state feedback, output feedback, filter and observer design. Singular perturbation methods also proved useful for the analysis of high-gain feedback systems and the
interpretation of other model order reduction techniques. More recently they have been applied to modeling and control of dynamic networks and certain classes of large-scale systems. This versatility of singular perturbation methods is due to their use of time-scale properties which are common to both linear and nonlinear dynamic systems.

The first survey [A4] of control theory applications of singular perturbations in 1976 included 130 references. The last six years have witnessed an even faster growth of this research area both in theoretical depth and breadth of applications, as evidenced by surveys and books [A1—A20]. The present survey of over 400 publications can only outline research directions with brief references to the representative issues and results. Each of the topics discussed here is rich enough for a detailed survey. This is particularly true of the references [B1—B2], which contribute to the continuing strengthening of mathematical foundations of classical and modern asymptotic methods. Let us only mention that a new coordinate-free formulation of singular perturbations and time scale properties appears in [B19], while [B13] and [B16] compare matched asymptotic expansions and averaging methods with multi-time scale formulations. Some basic results, proofs of correctness, and extensions of asymptotic methods have been obtained in [B2,B4,B6,B12,B15,B17,B18,B20]. Singular perturbation techniques are used to study bifurcations in ordinary differential equations [B11,B27], and jump phenomena in electrical circuits [B28]. Applications of distributions to the analysis of singularly perturbed systems appear in [B20,I24].

Our bibliography includes a group of references [01—043] on "singular systems." Under this heading we encompass seemingly diverse, but
closely related topics such as numerical methods for mixed differential-algebraic systems [01,03,032,041-043], descriptor systems [08,013,030,034,039], singular-singularly perturbed systems [02,05,010-019,025], generalized state (semi-state) equations for singular (degenerate) systems [06,020,023,024,027,029,033,035,036], etc. However, we do not attempt to review the intricate issues discussed in these references. The rich literature on partial differential equation methods, such as homogenization [A9,B9,B14], is also beyond the scope of this survey.

To make this text accessible to a broad audience of control engineers, Section 1 introduces singularly perturbed systems as a special class of two-time-scale systems. This section includes a survey of recent results on discrete-time systems. We then proceed with Sections 2 and 3 on system properties and linear composite control. Section 4 is dedicated to nonlinear and adaptive control, and Section 5 to stochastic filtering and control. A singular perturbation view of high-gain feedback systems is given in Section 6, which includes references on multivariable root loci and variable structure systems. Sections 7 and 8 deal with applications of time-scale methods to modeling and control of networks, Markov chains and other large scale systems with several control agents.

For ease of orientation the references are arranged in fifteen groups which largely correspond to sections or subsections of the text.
1. TWO-TIME-SCALE SYSTEMS

Examples of multi-time-scale systems abound [A4,M15] and include electrical networks [B28,K2,K3,K5,K7,K10-K14], power systems [K6,K8,K9, K15-K19], aircraft and rocket systems [H1,H4,H13,H15,H18-H20,H25,H27-H29, H31-H33], nuclear reactor systems [E14], scheduling systems [I25,L9], large space structures [E23], chemical kinetics [B24,G4], diffusion processes [I3,I10], population biology models [B10], and bifurcations [B11,B27,B29,B30]. Such systems can be modeled by the set of nonlinear differential equations

\[
\begin{align*}
\dot{x} &= f(x,z,t) \\
\dot{z} &= g(x,z,t)
\end{align*}
\]

where the n-dimensional vector \( x \) is predominantly slow and the m-vector \( z \) contains fast transients superimposed on a slowly varying "quasi-steady-state."

A linear time-invariant version, sometimes obtained by linearization of (1.1), takes the form

\[
\begin{bmatrix}
\dot{x} \\
\dot{z}
\end{bmatrix} =
\begin{bmatrix}
A & B \\
C & D
\end{bmatrix}
\begin{bmatrix}
x \\
z
\end{bmatrix}.
\]

The separation of states into those which are slow and those which are fast is a nontrivial modeling task [C24,K19,M15] demanding insight and ingenuity on the part of the analyst. In the absence of empirical estimates of \( \dot{x} \) and \( \dot{z} \), physical parameters such as time constants, loop gains, and inertias are examined to determine which states are slow and which are fast. As illustrated in [C17], a permutation and/or scaling of states is required
to obtain the state separable model (1.1) or (1.2). In some applications the slow and fast states of the overall system can be determined from the slow and fast states of each subsystem, but in general, the effect of interconnections on the speeds of some states may be by no means negligible. A new choice of state variables may then be needed to make the overall system model state separable. A time-scale modeling methodology which accomplishes this in networks and Markov chains is discussed in Section 7. If the linear model (1.2) satisfies (rather conservative) norm conditions \([C6]\) or conditions involving a correlation of states with eigenvalues \([K9, C20]\), its two-time-scale property is assured. Otherwise, a remodeling with a different selection of states is required.

1.1. Linear Systems

Linear time-invariant system (1.2) exhibits a two-time-scale property if it can be transformed into the block-triangular system

\[
\begin{bmatrix}
\dot{x} \\
\dot{\eta}
\end{bmatrix} =
\begin{bmatrix}
F_1 & B \\
0 & F_2
\end{bmatrix}
\begin{bmatrix}
x \\
\eta
\end{bmatrix}
\] (1.3)

such that the largest eigenvalue of \(F_1\) is smaller than the smallest eigenvalue of \(F_2\),

\[
\max_{i} |\lambda_i(F_1)| < \min_{j} |\lambda_j(F_2)| .
\] (1.4)

Application of the transformation \([C5,C6,C10,C11,C13,C14]\)

\[
\eta = z + Lx
\] (1.5)
where $L$ satisfies the algebraic Riccati equation

$$DL - LA +LBL - C = 0 \quad (1.6)$$

to (1.2) results in the block-triangular system (1.3) where

$$F_1 = A - BL, \quad F_2 = D + LB. \quad (1.7)$$

To completely separate the "slow" and the "fast" subsystems we let

$$\xi = x - M\eta \quad (1.8)$$

and choose $M$ to satisfy

$$(A - BL)M - M(D + LB) + B = 0. \quad (1.9)$$

If the eigenvalue separation (1.4) is sufficiently large, the achieved decomposition

$$\dot{\xi} = (A - BL)\xi \quad (1.10)$$
$$\dot{\eta} = (D + LB)\eta \quad (1.11)$$

defines the "slow" state $\xi$ and the "fast" state $\eta$ and relates them to the original variables $x$ and $z$. As we shall see, this relationship is made clearer when instead of the exact decomposition (1.10), (1.11), an approximate expression for $L$ can be obtained.

Approximations enter into the decomposition by way of solving (1.6) and (1.9) for $L$ and $M$ respectively, up to a specified degree of accuracy. In [C6,C11,C14] the determination of the matrices $L$ and $M$ is
iterative. When the separation of time-scales (1.4) is large, the convergence is quite rapid, typically a couple of iterations. Even when the separation (1.4) is small, as for example in the power system decomposition of [K9], the iterative scheme [C14] is still applicable, provided that the original choice of slow and fast states in (1.2) is valid. The iterations for solving (1.6) are related to the simultaneous subspace iterations [C8, C14]. An alternative scheme for separating the slow and fast subsystems are the quasi-steady-state iterations [C15, C17, K9], which remove the inconsistencies of the classical quasi-steady-state approach [K1].

With minor modifications, the iterative procedures for two-time-scale system decomposition apply to linear time-varying systems. A nonlinear version, appropriate for a class of nonlinear systems separable in x and z, is outlined in [C17]. Decomposition of multi-time-scale systems can also be achieved along the lines of repeated two-time-scale decompositions [K9].

That a singularly perturbed system

\[ \dot{x} = Ax + Bz \quad x(0) = x^0 \quad (1.12) \]
\[ \varepsilon \dot{z} = Cx + Dz \quad z(0) = z^0 \quad (1.13) \]

where \( \varepsilon > 0 \) is small and \( D^{-1} \) exists, is a two-time-scale system, can be easily seen by substituting \( C/\varepsilon \) and \( D/\varepsilon \) for C and D, respectively, in (1.2) to (1.11). Then it follows from (1.6) that an approximation of \( L \) is

\[ L = D^{-1}C + O(\varepsilon) \quad (1.14) \]

and hence the slow and the fast subsystems (1.10), (1.11) become
\[
\dot{\xi} = [A - BD^{-1}C + O(\varepsilon)]\xi \quad (1.15)
\]
\[
\varepsilon \ddot{\eta} = [D + O(\varepsilon)]\eta. \quad (1.16)
\]

This proves that the small eigenvalues of (1.12), (1.13) are close to the eigenvalues of \(A - BD^{-1}C\), while the large eigenvalues are close to those of \(\frac{1}{\varepsilon}D\). Hence, smaller values of \(\varepsilon\) result in wider separation of time scales.

If for \(\varepsilon = 0\) we denote \(\xi\) by \(x_s\), then (1.15) becomes the so-called reduced (or "quasi-steady-state") model of (1.12), (1.13),

\[
\dot{x}_s = (A - BD^{-1}C)x_s, \quad x_s(t_0) = x^0. \quad (1.17)
\]

Formally (1.17) can be obtained by setting \(\varepsilon \dot{x}_s = 0\) in (1.13) and substituting

\[
z_s = -D^{-1}Cx_s. \quad (1.18)
\]

From (1.5) and (1.14) we see that (1.18) is consistent with neglecting both \(\varepsilon\) and \(\eta\) in (1.5), that is, \(z_s\) is the "quasi-steady state" of \(z\) at \(\varepsilon = 0\). If we introduce the fast time scale

\[
\tau = \frac{t-t_0}{\varepsilon}, \quad \tau = 0 \text{ at } t = t_0, \quad (1.19)
\]

and denote \(\eta\) by \(z_f\) whenever we set \(\varepsilon = 0\), then (1.16) becomes the so-called boundary layer system

\[
\frac{dz_f(\tau)}{d\tau} = Dz_f(\tau), \quad z_f(0) = z^0 - z_s(t_0). \quad (1.20)
\]

We have thus shown that if \(D\) is a stability matrix, then a two-time-scale approximation of the state of (1.12), (1.13) is
\[ x(t) = x_s(t) + o(\varepsilon) \]  
\[ z(t) = z_s(t) + z_f(\tau) + o(\varepsilon) \]

where \( x_s(t) \) and \( z_f(\tau) \) are the limits as \( \varepsilon \to 0 \) of the exact slow and fast states \( \xi(t) \) and \( \eta(\tau) \), respectively. Based on (1.21), (1.22) a model order-reduction, related to modal methods [C2] and aggregation [C3], but specific to two-time-scale systems, is achieved by neglecting the fast subsystem while retaining the reduced model (1.17). The justification for this is that \( z_f(\tau) \) decays rapidly in an initial "boundary layer" interval after which the system response is essentially due to \( x_s(t) \), \( z_s(t) \). Because this order reduction is based on time-scale properties and not on linearity, it is also applicable to two-time-scale nonlinear systems.

1.2. Nonlinear Systems

The structure of the approximation (1.21) and (1.22) remains the same for singularly perturbed time-varying and nonlinear systems

\[ \frac{dx}{dt} = f(x,z,t) \quad x(t_o) = x^0 \]  
\[ \varepsilon \frac{dz}{dt} = g(x,z,t) \quad z(t_o) = z^0. \]

In the limit as \( \varepsilon \to 0 \), the asymptotically stable fast transient decays "instantaneously" leaving the reduced-order model in the t time-scale defined by the quasi-steady-states \( x_s(t) \) and \( z_s(t) \)
\[ \frac{dx}{dt} = f(x, z, t) \quad x_s(t_0) = x^0 \quad (1.25) \]

\[ 0 = g(x, z, t). \quad (1.26) \]

To obtain the fast parts of \( x \) and \( z \) we rewrite (1.23) and (1.24) in the fast time-scale \( \tau = \frac{t-t'}{\varepsilon} \), where \( t' \) is any fixed instant,

\[ \frac{dx}{d\tau} = \varepsilon f(x, z, t'+\varepsilon \tau) \quad (1.27) \]

\[ \frac{dz}{d\tau} = g(x, z, t'+\varepsilon \tau) \quad (1.28) \]

and again examine the limit as \( \varepsilon \to 0 \). Then \( dx/d\tau = 0 \), that is \( x = \) constant in the fast time-scale. The only fast variations are the deviations of \( z \) from its quasi-steady-state \( z_s \). Denoting them by \( z_f = z-z_s \) and letting \( \varepsilon = 0 \) in (1.27), (1.28), we obtain the fast subsystem

\[ \frac{dz_f}{d\tau} = g(x^0, z_s^0 + z_f(\tau), t_0), \quad z_f(0) = z^0 - z_s^0 \quad (1.29) \]

where the fixed instant \( t' \) has been chosen to be \( t_0 \) and \( x^0 \) and \( z_s^0 \) are fixed parameters. Conditions under which the slow subsystem (1.25), (1.26) and the fast subsystem (1.29) together yield the state approximation (1.21), (1.22) (and higher order approximations up to \( O(\varepsilon^k) \)) are given in [C1] and in more recent references surveyed in [A4]. These references assume that the fast transients described by (1.29) are asymptotically stable uniformly in \( x^0, z_s^0, t_0 \). For linear time-invariant systems this means that the large eigenvalues have large negative real parts so that the fast state rapidly reaches its quasi-steady-state \( z_s \). On the other hand,
mechanical and electromechanical systems often have lightly damped modes resulting in sustained high frequency oscillations. After linearization such systems can be put in the singularly perturbed form (1.12), (1.13) and a transformation similar to that of (1.5), (1.8) can be applied [K4]. However, in this case, the two-time-scale approximation (1.21), (1.22) is only valid up to a finite time. Under additional assumptions the slow phenomena in the oscillatory singularly perturbed system are adequately modeled by the slow subsystem which has an "averaging" effect on the fast oscillations. An example of this modeling procedure is the study of coherency in power systems where high frequency intermachine oscillations are not negligible [K4]. Related control studies of quasi-conservative large space structures are reported in [E32]. More general treatment of oscillatory two-time-scale systems and periodic structures is possible via averaging methods [A8,A14,B16] and homogenization [A9,B9,B14].

1.3. Discrete Time Systems

In recent years considerable progress has been made in formulation and analysis of two-time-scale discrete-time models [D2-D27]. The fact that the theory of difference equations is in most respects akin to that of ordinary differential equations [D1] suggests that a similar two-time-scale decomposition [D6] might be obtained for discrete-time systems. Attempts, however, to model general discrete-time systems with slow and fast modes in a strict singularly perturbed format encountered stability difficulties [D5,D8]. For a discretization of a continuous-time singularly perturbed system, the discretization interval can be compatible with either fast or slow time scales. The former case is considered in [D18]
where the discretization interval is $O(\epsilon)$. This case admits a hybrid two-time-scale approximation consisting of a discrete-time fast model and a continuous-time slow model. This decomposition has been applied to the design of linear optimal regulators in [D18]. A further development in [D27] relates this result with its continuous-time counterpart [E8]. The assumption that the interval is small or the number of time-steps is large is pursued by [D3,D10,D20] using asymptotic expansions. Other topics include the numerical solution of continuous-time boundary value problems [D7] and initial value problems [D15], discrete system initial value problems [D20] and boundary value problems [D22,D23], optimal control problems [D24, D25], limiting behavior of solutions of singular difference equations [D12], and multi-time methods for discrete systems [D9,D13].

The application of the block-diagonalization [C6] to two-time-scale discrete-time systems is straightforward. Paralleling the continuous time approach, references [D16,D19,D26] establish that the linear system

$$\begin{bmatrix} x(k+1) \\ z(k+1) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x(k) \\ z(k) \end{bmatrix}$$

and

$$\begin{bmatrix} x(0) \\ z(0) \end{bmatrix} = \begin{bmatrix} x^0 \\ z^0 \end{bmatrix}$$

exhibits a two-time-scale property if it can be transformed into the block-diagonal system

$$\begin{bmatrix} \xi(k+1) \\ \eta(k+1) \end{bmatrix} = \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix} \begin{bmatrix} \xi(k) \\ \eta(k) \end{bmatrix}$$

where the largest eigenvalue of $F_2$ is less than the smallest eigenvalue of $F_1$. Formally setting $z(k+1) = z(k)$ in (1.30) is equivalent to assuming that after
the decay of (stable) fast modes, the system (1.31) reaches the quasi-steady-state described by the slow subsystem

\[ x_s(k+1) = (A - BL_0)x_s(k), \quad L_0 = -(I_m - D)^{-1}C. \]  \hfill (1.32)

During the initial short transient period in which the fast modes are active, we assume that \( x_s(k) \) and \( z_s(k) \) are constant. Subtracting \( z_s(k) = -L_0x_s(k) \) from \( z(k) \) to obtain the fast part \( z_f(k) = z(k) - z_s(k) \), the fast subsystem of (1.30) can be expressed as

\[ z_f(k+1) = Dz_f(k), \quad z_f(0) = z^0 - z_s(0). \]  \hfill (1.33)

The singular perturbation parameter \( \epsilon \) can be introduced to explicitly express the magnitudes of system matrices as

\[ A = \hat{A}, \quad B = \epsilon^{1-\sigma}\hat{B}, \quad C = \epsilon\hat{C}, \quad D = \epsilon\hat{D} \]  \hfill (1.34)

where \( 0 \leq \sigma \leq 1 \) and \( \hat{A}, \hat{B}, \hat{C}, \) and \( \hat{D} \) are \( O(1) \). In this case \( BL_0 \) is \( O(\epsilon) \). Alternatively \( A, B, C, \) and \( D \) may be required to satisfy more general norm conditions [D19].
2. SYSTEM PROPERTIES

Attention is now focused on how certain properties of two-time scale systems; namely, stability, controllability, observability, and system zeros, can be deduced from lower-order subsystems in separate time scales.

2.1. Stability

In order to guarantee a stability property of the linear singularly perturbed system (1.12), (1.13), it is sufficient in view of the decoupled nature of subsystems (1.15), (1.16) that they both possess that property. More specifically, if the reduced system (1.17) and the boundary layer system (1.20) are asymptotically stable, then there exists an $\varepsilon^* > 0$ such that the original system (1.12), (1.13) is asymptotically stable for all $\varepsilon \in [0, \varepsilon^*]$, that is, for $\varepsilon$ sufficiently small. Such a result for time-varying and nonlinear systems dates from [G1]; see [A4]. For linear time-varying systems an explicit expression for the upper bound $\varepsilon^*$ is more recently presented in [C12]. A different condition via singular values appears in [G9]. In [K3] input-output stability is analyzed, while conditions for preservation of absolute stability in the presence of singular perturbations are derived in [G12, G17].

Considerable progress has been made in constructing Lyapunov functions for singularly perturbed systems [G2, G5, G11, G16]. For nonlinear systems which are linear in $z$,

\[
\dot{x} = f(x) + F(x)z \tag{2.1}
\]

\[
\varepsilon \dot{z} = g(x) + G(x)z \tag{2.2}
\]
where $G^{-1}(x)$ exists for all $x$, the Lyapunov function proposed in [G5] consists of two functions. The first function

$$v = a'(x)Q(x)a(x)$$ (2.3)

establishes the asymptotic stability of the slow subsystem $\dot{x}_s = a(x_s)$, where

$$a(x) = f(x) - F(x)G^{-1}(x)g(x),$$ (2.4)

and $Q(x) > 0$ satisfies

$$Q(x)a_x(x) + a_x'(x)Q(x) = -C(x), \quad a_x = \frac{3a}{3x}$$ (2.5)

for some differentiable $C(x) > 0$. The second function

$$w = (z + \Gamma g - P^{-1}\Gamma'F'v_x')'P(z + \Gamma g - P^{-1}\Gamma'F'v_x')$$ (2.6)

where $\Gamma = G^{-1}(x)$ and $P(x)$ satisfies

$$P(x)G(x) + G'(x)P = -I,$$ (2.7)

establishes the asymptotic stability (uniform in $x$) of the fast subsystem

$$\frac{dz_f}{dz} = G(x)z_f + g(x).$$ (2.8)

The Lyapunov function $V(x,z,\epsilon)$ for the complete system (2.1), (2.2) is

$$V(x,z,\epsilon) = v(x) + \frac{\epsilon}{2} w(x,z)$$ (2.9)

and can be used to estimate the dependence of the domain of attraction of $x = 0$, $z = 0$ on $\epsilon$. For example, the system
\begin{align*}
\dot{x} &= x - x^3 + z, \\
\epsilon \dot{z} &= -x - z
\end{align*}

is analyzed in [G5] using

\begin{equation}
V(x, z, \epsilon) = \frac{x^4}{4} + \frac{\epsilon}{4} (z + x + 2x^3)^2
\end{equation}

and for \( \epsilon < 0.01 \) the region includes \(|x| \leq 1, |z| \leq 10\), while for \( \epsilon < 0.005 \) the \( z \) bound is extended to \(|z| \leq 20\).

### 2.2. Controllability and Observability

Let us now consider the singularly perturbed control system

\begin{equation}
\begin{bmatrix}
\dot{x} \\
\epsilon \dot{z}
\end{bmatrix} =
\begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}
\begin{bmatrix}
x \\
z
\end{bmatrix} +
\begin{bmatrix}
B_1 \\
B_2
\end{bmatrix} u,
\begin{bmatrix}
x(0) \\
z(0)
\end{bmatrix} =
\begin{bmatrix}
x^0 \\
z^0
\end{bmatrix}
\end{equation}

\begin{equation}
y = [C_1 \ C_2]
\begin{bmatrix}
x \\
z
\end{bmatrix}
\quad x \in \mathbb{R}^n, \ z \in \mathbb{R}^m
\end{equation}

where \( u(t) \in \mathbb{R}^p \) is a control vector and \( y(t) \in \mathbb{R}^r \) is an output vector. Its slow and fast subsystems are

\begin{align*}
\dot{x}_s &= A_0 x_s + B_0 u, \quad x_s = x^0 \\
y_s &= C_0 x_s + D_0 u \\
\epsilon \dot{z}_f &= A_{22} z_f + B_2 u_f, \quad z_f(0) = z^0 - z_s(0) \\
y_f &= C_2 z_f
\end{align*}

where \( z_f = z - z_s, \ u_f = u - u_s, \ y_f = y - y_s \), and
The controllability conditions [J2]

\[
\begin{align*}
\text{rank}[\sigma I_n - A_o, B_o] &= n, \\
\text{rank}[\sigma I_m - A_{22}, B_2] &= m, \\
&\quad \forall \sigma \text{ complex (2.19)}
\end{align*}
\]

together imply the controllability of (2.12) for \( \varepsilon \) sufficiently small. The observability of the original system (2.12), (2.13) follows from observability of the subsystem pairs \((A_o, C_o)\) and \((A_{22}, C_2)\). Controllability results along these lines were introduced in [E5] for linear time-invariant systems and subsequently extended to linear time-varying systems in [G3] and [E25], and to the class of nonlinear systems in [G8]. Analogous results apply to continuous-time [E36] and discrete-time [D19] linear two-time-scale systems. In [E15] it is shown that the controllability of the singularly perturbed system (2.12) for \( \varepsilon > 0 \) does not necessarily require the controllability of the subsystems (2.14) and (2.16), although in practice, such weakly controllable systems are undesirable.

2.3. Analysis in the Frequency Domain

Taking one-sided Laplace transforms of (2.12), (2.13) yields, upon rearrangement,

\[
\begin{bmatrix}
A_{11} - \sigma I_n & A_{12} & B_1 \\
A_{21} / \varepsilon & A_{22} / \varepsilon - \sigma I_m & B_2 / \varepsilon \\
C_1 & C_2 & 0
\end{bmatrix}
\begin{bmatrix}
x(\sigma) \\
z(\sigma) \\
u(\sigma)
\end{bmatrix}
= \begin{bmatrix}
-x(0) \\
-z(0) \\
y(\sigma)
\end{bmatrix}.
\]

(2.20)
The matrix in (2.20) denoted by \( P(\sigma) \) is known as the system matrix [J2].

The invariant zeros [J7] of the system are those complex frequencies \( \sigma = \sigma_0 \) for which \( P(\sigma_0) \) loses rank and the corresponding invariant-zero directions are those vectors that lie in the null-space of \( P(\sigma_0) \). The asymptotic values of the invariant zeros and associated invariant-zero directions as \( \varepsilon \to 0 \) are the values computed from the system matrices

\[
\begin{align*}
P(s)(\sigma) &= \begin{bmatrix}
A_{11} - \sigma I_n & A_{12} & B_1 \\
A_{21} & A_{22} & B_2 \\
c_2 & c_2 & 0
\end{bmatrix}, & \quad P(f)(\sigma) &= \begin{bmatrix}
I_n & 0 & 0 \\
0 & A_{22} - \sigma I_m & B_2 \\
0 & c_2 & 0
\end{bmatrix}
\end{align*}
\]

(2.21)

associated with the slow and fast subsystems, respectively [J23].

Just as the system matrix \( P(\sigma) \) exhibits the internal structure associated with the singularly perturbed system (2.12), (2.13), the external relation between the input and output is described by \( y(\sigma) = G(\sigma)u(\sigma) \) where

\[
G(\sigma) = [C_1 \ C_2] \begin{bmatrix}
\sigma I_n - A_{11} & -A_{12} \\
-A_{21}/\varepsilon & \sigma I_m - A_{22}/\varepsilon
\end{bmatrix}^{-1} \begin{bmatrix}
B_2 \\
B_2/\varepsilon
\end{bmatrix}
\]

(2.22)

is an \( r \times p \) proper transfer-function matrix which is given to within an \( O(\varepsilon) \) as

\[
G(\sigma) = C_0(\sigma I_n - A_o)^{-1}B_o + C_2(\varepsilon \sigma I_m - A_{22})^{-1}B_2.
\]

(2.23)

As shown in [C7] the asymptotic forms of \( G(\sigma) \) for low and high frequencies are

\[
\begin{align*}
G_{\text{low}}(\sigma) &= C_0(\sigma I_n - A_o)^{-1}B_o + D_o & \quad \text{(2.24)} \\
G_{\text{high}}(\sigma) &= C_2(\varepsilon \sigma I_m - A_{22})^{-1}B_2
\end{align*}
\]

(2.25)
which, as expected, are the transfer-function matrices for the slow subsystem (2.14), (2.15) and the fast subsystem (2.16), (2.17), respectively. Moreover, the asymptotic values of the transmission zeros of \( G(\sigma) \) as \( \epsilon \to 0 \) are made up of those of \( G_{\text{low}}(\sigma) \) and \( G_{\text{high}}(\sigma) \) [J39].

Further insight into the frequency-domain analysis and design of singularly perturbed multivariable feedback systems is obtained by exhibiting the return-difference matrix at the input \([E2,E3]\), associated with the application of a linear state feedback control law. The static decoupling problem for two-time-scale multivariable systems is considered in [E18].
The decomposition of two-time scale systems into separate slow and fast subsystems suggests that separate slow and fast control laws be designed for each subsystem and then combined into a "composite" control of the original system. If the fast control is not needed, then only the slow "reduced" control is used. These ideas have motivated numerous two-time-scale designs [A4,E4-E9,E11-E14,E17-E42] which are now surveyed through a couple of representative problems.

3.1. Linear State Feedback

Suppose that the controls

\[ u_s = K_o s, \quad u_f = K_z f \]  

are separately designed for the slow and fast subsystems (2.14) and (2.16). In order to use

\[ u_c = u_s + u_f = K_o x_s + K_z z_f \]  

as a "composite" control for the full system (2.12), we must express the fictitious subsystem states \( x_s \) and \( z_f \) in terms of actual states \( x \) and \( z \).

Noticing that

\[ z_s = -A_{22}^{-1}(A_{21} + B_2 K_o) x_s \]  

and substituting the approximation \( x \approx x_s, \ z \approx z_s + z_f \) into (3.2) we obtain a realizable composite control.
When this control is applied to (2.12), the resulting feedback system is singularly perturbed. The time-scale decomposition shows that its slow and fast subsystem matrices are, respectively,

\[ u_c = K_0 x + K_2 \left[ z + A_{22}^{-1} (A_{21} + B_2 K_0) x \right]. \] (3.4)

where \( A_0 \) and \( B_0 \) are defined by (2.18). Therefore, \( K_0 \) and \( K_2 \) can be used for separate slow and fast eigenvalue placement, stabilization, or optimal state regulator designs provided that the pairs \( (A_0, B_0) \) and \( (A_{22}, B_2) \) are controllable or stabilizable. This approach was proposed in [E7,E8] and in [E12]. A proof of near-optimality of the composite control for a quadratic cost is given in [E8]. A two stage design extending these results to more general two-time-scale systems appears in [E36] and to linear time-varying systems in [E25]. When \( A_{22} \) is a stability matrix the fast subsystem need not be controlled. Then setting \( K_2 = 0 \) in (3.4) reduces \( u_c \) to the reduced order stabilizing control of [E4]. A similar slow-mode design was developed for the so-called linear multivariable tracking problem [E11,E22,E42]. The problem of eigenstructure assignment [E10,E16] is addressed in [E21].

Two-time-scale decomposition of near-optimal regulators for discrete-time linear systems appears in [D18,D25,D17]. The results in [D27] parallel the continuous time results of [E8]. Asymptotic expansions of the computationally attractive Chandrasekhar type of equations are presented in [E43]. Reference [E30] considers the case of designing a control with partial state
feedback to approximate the performance cost of a given full state feedback control where the given control need not be optimal. Extensions of the linear quadratic regulator to three-time-scale systems using two singular perturbation parameters are presented in [E20,E41], while [E40] considers near-optimum feedback with multiple time scales. Other work on the optimal regulator for linear singularly perturbed systems includes the asymptotic expansion methods of [E31], the near-optimum control of distributed parameter systems [E6], the near-optimum control of nuclear reactors with distributed parameters [E14], the control of quasi-conservative linear oscillatory systems [E32], the use of degeneracy to simplify the synthesis [E35], and the control of large space structures [E23].

3.2. Output Feedback and Observers

Static output feedback design of two-time-scale and singularly perturbed systems has been dealt with in several recent references [F7,F8, F11,F13,F14,F17]. To stress that the output feedback may be non-robust with respect to singular perturbations [F11], we consider

\[
\begin{align*}
\dot{x} &= z \\
\epsilon \dot{z} &= -x - z + u \\
y &= 2x + z.
\end{align*}
\]  

(3.6)

If instead of (3.6), its reduced order slow model

\[
\begin{align*}
\dot{x}_s &= -x_s + u \\
y_s &= x_s + u
\end{align*}
\]  

(3.7)
is used to design the output feedback $u = ky_s$ to place the eigenvalue at $-3$, then $k = 2$. Applying the same feedback law $u = 2y$ to the actual system (3.6) we see that the resulting feedback system

$$\dot{x} = z$$

$$\varepsilon \dot{z} = 3x + z$$

indeed has a slow eigenvalue $-3+0(\varepsilon)$ as desired. However, it also has an unstable fast eigenvalue $\frac{1}{\varepsilon}+0(1)$, that is, the design is nonrobust. The possibility of instability is due to the presence of $z$ in the output equation ("strong observability of parasitics"). It can be easily seen that a sufficient condition for a robust reduced order design is that $C_2 = 0$ and, hence, $D_o = 0$ in (2.15). Similar conditions in [F7,F8] also have the form of "weakly observable parasitics." If they are not satisfied, then a dynamic rather than static output feedback design should be used.

Typical dynamic feedback design is based on an observer [F1,F18]. We focus our attention on the reconstruction of the inaccessible state of the singularly perturbed system (2.12), (2.13) using the full-order observer

$$\begin{bmatrix}
\dot{x} \\
\dot{z}
\end{bmatrix} = (A - GC) \begin{bmatrix}
\hat{x} \\
\hat{z}
\end{bmatrix} + Gy + B_0 u$$

(3.9)

where $\hat{x}, \hat{z}$ is an estimate of the state $x, z$, and

$$A = \begin{bmatrix}
A_{11} & A_{12} \\
A_{21}/\varepsilon & A_{22}/\varepsilon
\end{bmatrix}, \quad B = \begin{bmatrix}
B_1 \\
B_2/\varepsilon
\end{bmatrix}.$$ 

(3.10)
The state reconstruction error $e(t)$ satisfies

$$
\dot{e}(t) = (A_e - GC)e(t), \quad e(0) = \left[ \begin{array}{c} \hat{x}(0) \\ \hat{z}(0) \end{array} \right] - \left[ \begin{array}{c} x(0) \\ z(0) \end{array} \right].
$$

(3.11)

Following the duality with composite control an observer gain matrix, composed of $G_o$ and $G_2$,

$$
G = \begin{bmatrix} A_{12}A_{22}^{-1}G_2 + G_0(I_m - C_2A_{22}^{-1}G_2) \\ G_2 \end{bmatrix}
$$

(3.12)

is designed in [F4-F6,F9,F10,F15]. It is pointed out in [F12] that although the observer gain (3.12) does not depend on $\varepsilon$, some knowledge of $\varepsilon$ is required in order to implement the observer. Nonrobustness of reduced order observers is analyzed in [F11,F12] and conditions are given under which a reduced-order observer can be designed to estimate the "slow" states of a singularly perturbed linear system from a knowledge of its reduced slow model. Decentralized identity observers for a large-scale system with two time-scales are treated in [F3] and the design of observer-based controllers for discrete two-time-scale systems is considered in [F17]. Another approach to reduced-order compensator design is described in [F19].
4. NONLINEAR AND ADAPTIVE CONTROL

The common theme of references surveyed in this section is that a scalar function (functional) defined for the full problem is decomposed into similar functions (functionals) for separate slow and fast subsystems. For optimal control problems this function is the Hamiltonian and for stabilization problems it is a Lyapunov function, which, in case of optimal regulators, coincides with the optimal value function.

4.1. Nonlinear Composite Control

Recent generalizations \[G16,G18\] of the composite control establish that if the boundary layer system (fast subsystem) and the reduced system (slow subsystem) are separately made uniformly asymptotically stable, then the full singularly perturbed system is stabilized for sufficiently small \(\varepsilon\). To avoid more complex notation we briefly outline an earlier result \([G6,G7,G11]\): a nonlinear infinite interval problem in which the system and cost to be optimized are

\[
\begin{align*}
\dot{x} &= a_1(x) + A_1(x)z + B_1(x)u, \quad x(0) = x^0 \\
\varepsilon \dot{z} &= a_2(x) + A_2(x)z + B_2(x)u, \quad z(0) = z^0 \\
J &= \int_0^\infty [p(x) + c'(x)z + z'Q(x)z + u'R(x)u]dt
\end{align*}
\]

subject to assumptions of differentiability, positivity and the existence of an equilibrium at \(x = 0, z = 0\). This problem is difficult even for well-behaved nonlinear systems. The presence of \(1/\varepsilon\) terms in the Hamilton-Jacobi
equation increases the difficulties. However, avoiding the full problem, and taking advantage of the fact that as $\varepsilon \to 0$ the slow and the fast phenomena separate, we can define two separate lower-dimensional subproblems. The slow subproblem is to optimally control the slow subsystem

$$\dot{x}_s = a_o(x_s) + B_o(x_s)u_s, \quad x_s(0) = x^0$$

(4.4)

with respect to

$$J_s = \int_0^\infty [p_o(x_s) + 2\sigma'(x_s)u_s + u_s' R(x_s) u_s] dt$$

(4.5)

where

$$a_o = A_1 A_2^{-1} a_2, \quad \sigma_o = B_2 A_2^{-1} (Q A_2^{-1} a_2 - \frac{1}{2} \sigma),$$

$$B_o = B_1 - A_1 A_2^{-1} B_2, \quad R_o = R + B_2' A_2^{-1} Q A_2^{-1} B_2,$$

$$p_o = p = \sigma' A_2^{-1} a_2 + a_2' A_2^{-1} Q A_2^{-1} a_2.$$ (4.6)

Assuming the existence of the optimal value function $L(x_s)$ satisfying the lower dimensional Hamilton-Jacobi equation

$$0 = (p_o - s R^{-1} \sigma_o) + L_x (a_o - B_o R^{-1} \sigma_o) - \frac{1}{2} L_x B_o R^{-1} B_o' L'_x, \quad L(0) = 0$$

(4.7)

the minimizing control for (4.4), (4.5) is

$$u_s = -R_o^{-1} (\sigma_o + \frac{1}{2} B_o' L'_x)$$

(4.8)

where $L_x$ denotes the derivative of $L$ with respect to its argument $x_s$. Then $L(x_s)$ is also a Lyapunov function which guarantees the asymptotic stability of $x_s = 0$ for (4.4) controlled by $u_s$. 
\[ \dot{x}_s = a_o - B_o R_o^{-1} \left( \sigma_o + \frac{1}{2} B_o' L_o' \right) = \dot{a}_o(x_s) \]  

(4.9)

and provides an estimate \( D \) of the region of attraction of \( x_s = 0 \).

Introducing \( z_f = z - z_s \), \( u_f = u - u_s \) and neglecting \( O(\varepsilon) \) terms, we define the fast subproblem in the \( \tau \)-scale (1.19) as

\[ \frac{d z_f}{d \tau} = A_2(x) z_f + B_2(x) u_f, \quad z_f(0) = z^o - z_s(0), \]  

(4.10)

\[ J_f = \int_0^{\infty} (z_f' Q(x) z_f + u_f' R(x) u_f) d\tau, \]  

(4.11)

which is to be solved for every fixed \( x \in D \). It has the familiar linear quadratic form and under a controllability assumption its solution is

\[ u_f(z_f,x) = -R_1^{-1}(x) B_2^t(x) K_f(x) z_f \]  

(4.12)

where \( K_f(x) \) is the positive-definite solution of the \( x \)-dependent Riccati equation

\[ 0 = K_f A_2 + A_2^t K_f - K_f B_2 R_1^{-1} B_2^t K_f + Q. \]  

(4.13)

The control (4.12) is stabilizing in the sense that the fast feedback system

\[ \frac{d z_f}{d \tau} = (A_2^o - B_2 R_1^{-1} B_2^t K_f) z_f = \ddot{A}_2(x) z_f \]  

(4.14)

has the property that \( \text{Re} \lambda[\ddot{A}_2(x)] < 0 \), \( \forall x \in D \).

The design is completed by forming the composite control \( u_c = u_s + u_f \), in which \( x_s \) is replaced by \( x \) and \( z_f \) by \( z + A_2^{-1}(a_2 - B_2 u_s(x)) \), that is,

\[ u_c(x,z) = u_s(x) - R_1^{-1} B_2^t K_f (z + A_2^{-1}(a_2 - B_2 u_s(x))) \]

\[ = -R_o^{-1} \left( \sigma_o + \frac{1}{2} B_o' L_o' \right) - R_1^{-1} B_2^t K_f (z + \ddot{A}_2^{-1} a_2) \]  

(4.15)
where

\[
\bar{a}_2(x) = a_2 - \frac{1}{2} B_2 R^{-1}(b_{1}' L_1 x + b_{2}' V_1), \quad \bar{a}_2(0) = 0,
\]

\[
V_1' = -(\sigma' + 2a_2 K + L_1 A_2) \bar{a}_2^{-1} \quad A_1 = A_1 - B_1 R^{-1} B_2 K.
\]

Under the conditions stated in [Gil], there exists \( \varepsilon^* \) such that \( \forall \varepsilon \in (0, \varepsilon^*) \), the composite control \( u_c \) stabilizes the equilibrium \( x = 0, z = 0 \) of the full system (4.1), (4.2). An estimate of the region of attraction is also given. The corresponding cost \( J_c \) is bounded and it is near-optimal in the sense that \( J_c \rightarrow J_s \) as \( \varepsilon \rightarrow 0 \). Thus the problem (4.1), (4.2), (4.3) is well posed with respect to \( \varepsilon \).

In [Gil] a composite Lyapunov function of the type (2.3), (2.6), (2.9) is used. More general Lyapunov functions in [Gl6] remove the restriction of the linearity with respect to \( z \).

4.2. Trajectory Optimization

For the trajectory optimization problem the control interval is finite, \( t \in [t_0, T] \), and some conditions are imposed on \( x, z \) of the full system (1.23), (1.24) at both ends of the interval. A control \( u(t) \) is sought to minimize

\[
J = \int_{t_0}^{T} V(x, z, u, t) dt.
\]

The necessary optimality condition consists of \( \frac{\partial H}{\partial u} = 0 \) and
which is a $(2n+2m)$-dimensional singularly perturbed boundary value problem, where $H$ is the Hamiltonian and $p$ and $\varepsilon q$ are the adjoint variables. Typical optimal trajectory consists of a slow "outer" part with "boundary layers" at the ends. In the limit as $\varepsilon \to 0$ the problem decomposes into one slow and two fast subproblems. The slow ("outer") subproblem

\[
\dot{z}_s = -\frac{\partial H_s}{\partial p_s}, \quad \dot{p}_s = -\frac{\partial H_s}{\partial z_s}
\]  

is $2n$-dimensional. To satisfy the remaining $2m$ boundary conditions, the layer ("inner") corrections $z_{fL}(\tau_L)$, $z_{fR}(\tau_R)$ for $z$, and $q_{fL}(\tau_L)$, $q_{fR}(\tau_R)$ for $q$ are determined from the initial (L) and final (R) boundary layer systems

\[
\frac{dz_{fL}}{d\tau_L} = \frac{\partial H_{fL}}{\partial q_{fL}}, \quad \frac{dq_{fL}}{d\tau_L} = -\frac{\partial H_{fL}}{\partial z_{fL}}
\]

\[
\frac{dz_{fR}}{d\tau_R} = \frac{\partial H_{fR}}{\partial q_{fR}}, \quad \frac{dq_{fR}}{d\tau_R} = -\frac{\partial H_{fR}}{\partial z_{fR}}
\]

where $\tau_L$ is the same as $\tau$ in (1.19), while $\tau_R = \frac{T-t}{\varepsilon}$ is the reversed fast time scale. The results of these subproblems are used to form approximations of the type (1.21), (1.22). Analytical issues of such approximations have been discussed in [H3,H5-H12,H17,H21,H24,H26,H30], and for the time optimal control problem in [H16,H22]. Conditions for asymptotic validity
are given in [H6,H7,H12,H30]. A trajectory optimization problem for systems of the type (4.1), (4.2) is solved in [H21].

In flight dynamics both low thrust (aircraft) and high thrust (missile) conditions can be modeled as singular perturbations. Several applications to specific problems arising in jet engine control, missile guidance and energy management have been reported in [H1,H2,H4, H13,H15,H18,H19,H22,H25,H27-H33]. Some approaches, such as [H20], use sequential multi-time-scale modeling and obtain solutions in convenient feedback form. Problems of scaling are addressed in [H27,H28,H32,H33] and indicate that more research is required to properly introduce singular perturbation parameters for specific aircraft and missile control problems. These issues are more complicated in the presence of singular arcs [H24].

4.3. Adaptive Control

The study of robustness of model-reference adaptive schemes in the presence of singular perturbations has recently been initiated. The general formulation examines situations when the order of the model is equal to the order of the slow part of the unknown plant and the model-plant "mismatch" is due to the fast part of the plant. A fundamental requirement for feasibility of an adaptive scheme is that it be robust, that is tolerate a certain model-plant mismatch. The singular perturbation parameter $\epsilon$ is a convenient parameterization of this mismatch.

The robustness of continuous-time identifiers and observers when the parasitics are weakly observable is established in [G14,G19]. The bounds for output/parameter error indicate possibilities for reducing the
error by a proper choice of the input signal. When the parasitics are
strongly observable, the problem is no longer robust. Robustness is
re-established either by using a low-pass filter at the output, or by an
appropriate modification of the adaptive laws [G15]. Analogous results for
discrete time parallel and series-parallel identifiers have been obtained in
[G13,G21]. The methodology in [G14,G19] has been to transform the singularly
perturbed plant (2.12) into

\[ \dot{x} = A_0 x + B_0 u + A_{12} \zeta \]  
(4.23)

\[ \dot{\zeta} = A_f \zeta + \varepsilon A_f^{-1} B_f \dot{u} \]  
(4.24)

where

\[ \zeta = z + L x + A_f^{-1} B_f u \]  
(4.25)

and \( L \) is the same matrix as in Section 1.1. The matrices \( A_0, B_0 \) of the
dominant part (4.23) are matchable by the model, while the parasitic part
(4.24) introduces a disturbance \( A_{12} \zeta \). It is crucial to observe that the
disturbance is persistently excited by the derivative of the input signal
\( \dot{u}(t) \) and is large if \( u(t) \) has high frequencies in the parasitic range. For
this reason the input signal should achieve its richness condition at
dominant, rather than parasitic frequencies. A significant conclusion is
that, contrary to common belief, white noise and similar wide-band signals
are not as appropriate as some lower frequency signals.

The situation is more complex in adaptive control where the plant
input depends on the adaptive feedback. In this case the parasitics can
destroy global stability and boundedness properties. The singular perturba-
tion approach in [G20,G23] has led to a modified adaptation law which
guarantees the existence of a region of attraction from which all signals converge to a residual set whose size depends on disturbances and the mismatch parameter $\varepsilon$. The modified law achieves robustness by sacrificing the ideal convergence in the parasitic-free case, i.e., the residual set is not zero even if $\varepsilon = 0$. It is of interest that the development of the modified law uses a two-time-scale Lyapunov function of the type discussed in subsection 2.1. The details are given in a recent monograph [G23].
5. STOCHASTIC FILTERING AND CONTROL

Research in singular perturbation of filtering and stochastic control problems with white noise inputs has revealed difficulties not present in deterministic problems. This is due to the fact that the input white noise process "fluctuates" faster than the fast dynamic variables, no matter how small \( \varepsilon \) is. In the limit as \( \varepsilon \to 0 \), the fast variables themselves tend to white noise processes, thus losing their significance as physically meaningful dynamic variables. The papers dealing with stochastic differential equations and diffusion models such as [13,17,19,122,126] have also indicated the importance of attaching clear probabilistic meaning to time scales.

5.1. LQG Problems

To illustrate the problems arising in the singularly perturbed formulation of systems with white noise inputs, consider the linear system

\[
\dot{x} = A_{11} x + A_{12} z + G_1 w \quad (5.1a)
\]

\[
\varepsilon \dot{z} = A_{21} x + A_{22} z + G_2 w \quad (5.1b)
\]

where \( w(t) \) is white Gaussian noise. An approximation to (5.1) is desired that is valid for small \( \varepsilon \) and is simpler than (5.1). Setting \( \varepsilon = 0 \) in (5.1) is inadequate, since

\[
\bar{z} = -A_{22}^{-1} (A_{21} \bar{x} + G_2 w) \quad (5.2)
\]
has a white noise component and, therefore, has infinite variance. Although (5.2) may be substituted for $z$ in defining a reduced (slow) subsystem, $	ilde{z}$ cannot serve as an approximation for $z$ in the mean square sense. Under the assumption

\[ \text{Re } \lambda(A_{22}) < 0, \quad \text{Re } \lambda(A_{11} - A_{12}A_{22}^{-1}A_{21}) < 0 \]  

(5.3)

the mean-square convergence

\[
\lim_{\varepsilon \to 0} E\{(x - x_d)(x - x_d')\} = 0
\]
\[
\lim_{\varepsilon \to 0} E\{(z - z_d)(z - z_d')\} = 0
\]  

(5.4)

was demonstrated in [II, III] for $x_d$, $z_d$ defined by

\[
\dot{x}_d = (A_{11} - A_{12}A_{22}^{-1}A_{21})x_d + (G_1 - A_{12}A_{22}^{-1}G_2)w
\]
\[
\varepsilon \dot{z}_d = A_{21}x_d + A_{22}z_d + G_2w.
\]  

(5.5)

For the linear filtering of (5.1) with respect to the observations

\[ y = C_1x + C_2z + v \]  

(5.6)

where $v(t)$ is a white Gaussian noise independent of the process noise $w(t)$, the analysis in [II] demonstrates that the Kalman filter can be approximately decomposed into two filters in different time scales, thereby yielding estimates of the slow and fast states. Similar results are obtained for near-optimal smoothing [III], and state estimation with uncertain singular perturbation parameter $\varepsilon$ in (5.1) [II9].
For the singularly perturbed LQG control problem

\[ \dot{x} = A_{11}x + A_{12}z + B_1u + G_1w \]

\[ \varepsilon \dot{z} = A_{21}x + A_{22}z + B_2u + G_2w \]  \hspace{1cm} (5.7)

\[ y = C_1x + C_2z + v \]  \hspace{1cm} (5.8)

\[ J = E\{x'(T)\Gamma_1x(T) + 2\varepsilon x'(T)\Gamma_{12}z(T) + \varepsilon z'(T)\Gamma_2z(T) \]

\[ + \int_0^T (x'L_1x + 2\varepsilon x'L_{12}z + z'L_2z + u'Ru) dt \} \]  \hspace{1cm} (5.9)

it was demonstrated in [18] that the optimal solution to (5.7)-(5.9) may be approximately obtained from the solutions of two reduced order stochastic control problems in the slow and fast time scales. The drawback of the formulation (5.7)-(5.9) is that it results in J being \( O(1/\varepsilon) \). To avoid divergent performance indices, it is required that \( L_2 = O(\varepsilon), \Gamma_2 = O(\varepsilon^r) \).

A new approach [129] to this problem is based on singularly perturbed Lyapunov equations. The two-time scale solutions yield \( O(\varepsilon^r) \) \((r \geq 1)\) approximation to the optimal performance. For \( r=1 \), the results of [129] are equivalent to those of [18].

A conclusion from [11,18,129] is that in order to obtain well-defined reduced-order filtering and control problems as \( \varepsilon \to 0 \), it is necessary to scale the contribution of the white noise disturbances to the fast subsystem in some sense. Reference [117] introduces small parameters as powers of \( \varepsilon \) as follows
\[ \dot{x} = A_{11}x + A_{12}z + B_1u + G_1w \]
\[ \varepsilon \dot{z} = \varepsilon^\beta A_{21}x + A_{22}z + B_2u + \varepsilon^\alpha G_2w \]  \hspace{2cm} (5.10)
\[ y_1 = C_{11}x + C_{12}z + v_1 \]
\[ y_2 = \varepsilon^\nu C_{21}x + C_{22}z + \varepsilon^\nu v_2 \]  \hspace{2cm} (5.11)

\[ J = E \{ x'(T)\Gamma_1x(T) + 2\varepsilon x'(T)\Gamma_{12}z(T) + \varepsilon z'(T)\Gamma_2z(T) \]
\[ + \int_0^T (x'\Lambda_1x + 2\varepsilon^\delta x'\Lambda_{12}z + \varepsilon^2 z'\Lambda_2z + u'Ru) \, dt \}. \]  \hspace{2cm} (5.12)

The parameters \( \alpha, \beta, \nu, \delta \) represent the relative size of the small parameters within the system, with respect to the small time constants of the fast subsystem. The inclusion of a separate observation channel \( y_2 \) for the fast subsystem is essential, since for \( \alpha > 0 \), the fast variables cannot be estimated in a meaningful manner from the slow observation channel (signal-to-noise ratio tends to zero). The analysis of [117] shows that the performance index is finite if

\[ \delta \geq \left( \frac{1}{2} - \alpha \right). \]  \hspace{2cm} (5.13)

Furthermore, a well-defined formulation also requires that

\[ 0 \leq \alpha = \nu \leq \beta \leq \frac{1}{2}. \]  \hspace{2cm} (5.14)

The restriction \( \alpha = \nu \) is crucial; otherwise, either the fast variables are not observed due to noisy observations \( (\alpha > \nu) \), or they are observed noiselessly \( (\alpha < \nu) \) in the limit as \( \varepsilon \to 0 \). If \( \alpha > \frac{1}{2} \), the problem becomes deterministic as \( \varepsilon \to 0 \), and if \( \beta > \frac{1}{2} \), the coupling between \( x \) and \( z \) becomes of
order less than $O(\varepsilon^{1/2})$. The constraint $\beta \geq \alpha$ insures that the state $z$ is predominantly fast, and relaxing it causes no conceptual difficulties.

Setting $\alpha = \beta = \nu = 0$, yields the results of [I8]. In this case $\delta = \frac{1}{2}$ to yield a finite performance index. Setting $\alpha = \nu = \frac{1}{2}$ and $\delta = 0$ results in the full weighting of the fast variable. This important case, which is the only one to yield a well-defined stochastic process in the fast time scale, has been studied separately in [I27] and [I18]. They have established the weak convergence, as $\varepsilon \to 0$, of the fast stochastic variable $z$ which satisfies the Ito equation

$$\varepsilon dz = Az \, dt + \sqrt{\varepsilon} \, Gd\omega; \quad \text{Re} \lambda(A) < 0$$ (5.15)

where $\dot{\omega}(t)$ is Gaussian white noise with covariance $W$. The results of [I18, M27] show that

$$\lim_{\varepsilon \to 0} z(t; \varepsilon) = \tilde{z} \quad \text{weakly}$$ (5.16)

where $\tilde{z}$ is a constant Gaussian random vector with covariance $P$ satisfying the Lyapunov equation

$$AP + PA' + GWG' = 0.$$ (5.17)

Alternative formulations of the linear stochastic regulator problem have been reported in [I2, I6, I10, I16, I20]. Colored noise disturbance in the fast subsystem is assumed in [I16], which accounts for situations when the correlation time of the input stochastic process is longer than the time constants of fast variables. When the fast subsystem is stable, the results of [I16] demonstrate that the optimal solution to the stochastic
regulator problem can be approximated by the optimal solution of the slow subproblem, and the performance index does not diverge.

5.2. Nonlinear Stochastic Control

A composite control approach for a class of nonlinear systems driven by white noise disturbances appears in [123] as a stochastic version of [G6,G7,G11], namely

\[
\begin{align*}
\dot{x} &= (c(x)z + d(x) + 2\beta(x)u)dt + \sqrt{2} \, dw_1 \\
\epsilon \dot{z} &= (a(x)z + b(x) + 2\alpha(x)u)dt + \epsilon \sqrt{2} \, dw_2
\end{align*}
\]

(5.18)

\[
J^\epsilon_{x,z}(u(\cdot)) = E \int_0^\infty e^{-\nu t} \left[ (f(x) + h(x)z)^2 + u^2 \right] dt
\]

(5.19)

where \(w_1(t), w_2(t)\) are standard Wiener processes independent of each other.

The optimal feedback law for (5.18), (5.19) is given by

\[
u^\epsilon(x,z) = -\beta(x)V^\epsilon_x(x,z) - \frac{\alpha(x)V^\epsilon_z(x,z)}{\epsilon}
\]

(5.20)

where \(V^\epsilon(x,z)\) is the Bellman function

\[
V^\epsilon(x,z) = \inf_{u(\cdot)} J^\epsilon_{x,z}(u(\cdot))
\]

(5.21)

As \(\epsilon \to 0\), the optimal solution (5.20), (5.21) converges to the solutions of two reduced-order problems, which yield an \(O(\epsilon)\) approximation to the optimal performance.
The slow control problem is
\[
dx = \left( -\frac{c}{a} (b + 2au_s) + d + 2bu_s \right) dt + \sqrt{2} dw
\]
\[
J^O_x(u_s(\cdot)) = E \int_0^\infty e^{-\nu t} \left[ (f - \frac{h}{a} (b + 2au_s))^2 + y_s^2 \right] dt. \tag{5.22}
\]

The fast problem is an x-dependent deterministic optimal control problem given by
\[
\varepsilon \dot{z}_f = az_f + 2au_f
\]
\[
J^O_z(u_f(\cdot)) = \int_0^\infty (h z_f^2 + u_f^2) dt. \tag{5.23}
\]

The composite control is formed as
\[
u_c(x,z) = u_s(x) + u_f(x,z)
\]
where \(u_s(x)\) is the optimal control of (5.22) and \(u_f(x,z)\) is the optimal control of (5.23).

Under mild regularity conditions, the existence of \(u_s(x)\) has been established. This is in contrast to the deterministic problem of [G6, G7, G11], where the existence of the slow optimal control had to be assumed. Furthermore, the composite feedback control \(u_c(x,z)\) maintains the payoff bounded as \(\varepsilon \to 0\) for \(\nu\) sufficiently large. Also
\[
V^c(x,z) \to V^O_s(x) \text{ pointwise as } \varepsilon \to 0 \tag{5.25}
\]
where \(V^O_s\) is the Bellman function of the slow subproblem (5.22).
Noticing that [I23] has scaled the white noise input to the fast subsystem by the factor $\varepsilon\sqrt{2}$, we understand why the fast optimal control problem becomes deterministic in the limit $\varepsilon \to 0$. A more realistic formulation may be obtained by using the scaling $\sqrt{2\varepsilon}$.

Singular perturbations of quasi-variational inequalities arising in optimal stochastic scheduling problems are investigated in [I25]. The system considered is the diffusion process

$$dx_t = m(x_t)dt + \frac{\sqrt{2\varepsilon}}{\varepsilon^2} \sigma(x_t)dw_t. \quad (5.26)$$

With this process we associate an index process $i_t$, a pair of cost rates $\{f^i(x), i=0,1\}$, and a pair of switching costs $\{k_{ij}(x); i,j=0,1; i\neq j\}$. Scheduling the system (5.26) consists of changing the cost rate $f^i(x)$ to a more favorable rate $f^j(x)$ at discrete times $S_m$. A cost $k_{ij}(x)$ is incurred for each change. Changes are based on the full observation of $(x_t,i_t)$ which is a Markov process. The solution of the optimal policy $\{S_m^i, i=0 \ldots \infty, m=0 \}$ with respect to a certain cost function leads to a system of second-order singularly perturbed quasi-variational inequalities. Asymptotic approximations to the optimal solution are obtained using standard singular perturbation analysis of boundary value problems.

5.3. Wide-Band Input Noise Formulations

Keeping in mind the limiting behavior of the fast variable, it is perhaps more realistic to model the input disturbances as wide-band noise [I14]. In formulations for linear systems [I15], nonlinear systems
[I12], and linear filtering problems [I4], both the parasitic elements of the system and the correlation time of the input stochastic process have been represented by the same perturbation parameter. By doing so, information about the relative behavior of the two asymptotic phenomena is lost. In fact, it has been shown in [I18] that the interaction between the two idealization procedures does affect the reduced order model. The following second order system is considered in [I18],

$$\varepsilon \ddot{x} + \dot{x} = a(x) + b(x)\nu^\mu$$

(5.27)

where \(\nu^\mu\) is exponentially-correlated noise with correlation time \(\mu\). It is shown that for sufficiently small \(\varepsilon\) and \(\mu\), \(x(t)\) can be modeled as a Markov process \(\bar{x}(t)\), which is the solution of the Ito equation

$$d\bar{x} = [a(\bar{x}) + \frac{0.5}{1+\varepsilon/\mu} \frac{\partial b}{\partial \bar{x}} b(\bar{x})S(0)]dt + d\omega$$

(5.28)

where \(S(\cdot)\) is the spectrum of \(V\). The important feature of (5.28) is the dependence of the drift coefficient on the ratio \(\varepsilon/\mu\). This has been generalized in [I28] to the nonlinear singularly perturbed system

$$\dot{x} = a_1(x) + A_{12}(x)z + B_1(x)\nu^\mu$$

$$\varepsilon \dot{z} = a_{21}(x) + A_2z + B_2(x)\nu^\mu$$

(5.29)

where \(\nu^\mu(t)\) is a wide-band zero mean stationary process with correlation matrix

$$E(\nu^\mu(t)(\nu^\mu(t+\tau))) = \frac{1}{\mu} R(\tau).$$

(5.30)
Assuming that Re $\lambda(A_2) < 0$, a diffusion model for $\tilde{x}(t)$ is obtained where the drift coefficient depends on $\varepsilon/\mu$. Similar scaling problems appear in bifurcation of stochastic differential equations.[B30].
6. HIGH-GAIN FEEDBACK SYSTEMS

Multivariable generalizations of classical single-input single-output high-gain feedback arise in a number of control problems: disturbance rejection, parameter uncertainty, decoupling of large-scale systems, and are implicit in cheap control problems [J8,J12,J16–J18,J20,J36,J41,J43,J48, J50,J52,J59,J60], almost singular state estimation [J45,J61], and variable structure systems with sliding mode [A3,A6,A7,J14,J26–J30,J58]. High gain feedback is treated in the context of multivariable root loci [J5,J21,J22, J33,J37,J42,J61,J62] and multivariable optimal root loci [J6,J19,J25,J38], and, more recently, the theory of invariant subspaces [J46,J53]. There is a synergism between these approaches and singular perturbations. Singular perturbation methods provide a unifying framework for the analysis and elucidation of diverse high-gain feedback problems in both linear and non-linear systems. Conversely, high-gain systems, which seldom appear in the standard form (1.23),(1.24), motivate the investigation of more general forms of singularly perturbed systems.

6.1. High-Gain and Time Scales

Perhaps the simplest multivariable high-gain feedback system is

\[ \dot{x} = Ax + Bu, \quad x \in \mathbb{R}^n, \quad u \in \mathbb{R}^p \]  \hspace{1cm} (6.1)

\[ u = \frac{1}{\varepsilon} Cx = \frac{1}{\varepsilon} y, \quad y \in \mathbb{R}^p. \] \hspace{1cm} (6.2)

Substitution of (6.2) into (6.1) yields

\[ \varepsilon \dot{y} = CB \dot{y} + \varepsilon CAx \] \hspace{1cm} (6.3)
and shows that, if $CB$ is nonsingular, then $y$ is the fast variable. After the fast transient near the range space of $B$ decays, the slow motion is continuous near the null space of $C$, [J14]. From the theory of asymptotic root loci [J5, J22, J33, J42] as $\epsilon \to 0$ the $p$ fast eigenvalues of the high-gain system (6.1), (6.2) tend to infinity (the infinite zeros) along the asymptotes defined by the directions of $\lambda_i(CB)$, while the $n-p$ slow eigenvalues tend to the transmission zeros (finite zeros) of the open-loop system (6.1) with the output $y = Cx$. Thus when $CB$ is nonsingular system (6.1), (6.2) can be expressed in the standard form (1.12), (1.13) and analyzed using the methods of Section 1.

When $CB$ is singular the situation is more complicated. For example, under the condition

$$CA^iB = 0, \quad i=0,1,\ldots,q-2$$

(6.4)

$$CA^{q-1}B$$

is nonsingular,

which is sufficient for the invertibility of (6.1) with $y = Cx_0$ [J1] (and corresponds to case $q$ in cheap control and singular arc problems [J8, J12, J52], that is to infinite poles of uniform order $q$ in multivariable root locus problems [J37]), there are $q$ fast time scales

$$t-t_0, \quad t-t_0^2, \quad \ldots, \quad t-t_0^q$$

(6.5)

As $\epsilon$ tends to zero, the slow eigenvalues coincide with the system transmission zeros as before, while the fast eigenvalues form the familiar Butterworth pattern of asymptotic root locus theory [J5, J22, J33, J37, J42].
6.2. Composite State Feedback and Cheap Control

Let matrix C be free to be chosen in a state feedback design and, without loss of generality, consider that (6.1) is in the form

\[ \begin{align*}
\dot{x}_1 &= A_{11}x_1 + A_{12}x_2, \quad x_1 \in \mathbb{R}^{n-p}, \quad x_2 \in \mathbb{R}^p \\
\dot{x}_2 &= A_{21}x_1 + A_{22}x_2 + B_2u, \quad u \in \mathbb{R}^p
\end{align*} \]  

and \( B_2 \) nonsingular. Then the composite high-gain feedback

\[ u = \frac{1}{\varepsilon} cx = \frac{1}{\varepsilon} [K_f K_1 x_1 + K_f x_2] \]  

(6.8)

can be used for a separate assignment [J14] of the slow and fast eigenvalues to locations

\[ \lambda_j (A_{11} - A_{12} K_s) + O(\varepsilon), \quad j = 1, \ldots, p \]  

(6.9)

\[ \frac{1}{\varepsilon} [\lambda_j (B_2 K_f) + O(\varepsilon)], \quad j = p+1, \ldots, n. \]  

(6.10)

Developments of this approach for high-gain state feedback and error-actuated tracking controllers are reported in [J24, J31, J34, J35, J44, J51]. Geometric conditions whereby the multivariable high-gain feedback system is insensitive to disturbances and (possibly large) parameter variations are explored in [J54]. An analysis of the interaction of actuator and sensor parasitics (frequently neglected in feedback systems design) with the fast system modes is undertaken in [J56].

The slow and fast gains can alternatively be determined from the solution of two separate lower-order regulator problems so as to provide a near-optimal composite state feedback control for the original system (6.1).
A cheap control problem analogous to the preceding eigenvalue assignment problem arises when the system (6.1) is regulated with respect to a quadratic performance index having small (cheap) penalty on $u$

$$J = \frac{1}{2} \int_0^\infty [x'Qx + \epsilon^2 u'Ru]dt \quad (6.11)$$

and $\epsilon$ is a small positive parameter. Assuming $B'QB > 0$, a composite control [J14] of the form (6.8), where $K_s$ and $K_f$ are the state feedback gain matrices associated with the respective solutions of slow and fast regulator problems, is near-optimal in the sense that the performance is $O(\epsilon^2)$ close to the optimum cheap performance. As $\epsilon \to 0$, the original system (6.1) under cheap control exhibits a two-time-scale response in that an initial fast transient (boundary layer) is followed by a slow motion on the singular arc determined by a dynamical system and matrix Riccati equation of reduced order $n-p$ [J8,J12,J32]. In the limit $\epsilon = 0$, the stable fast modes decay instantaneously, so that the optimal solution is given by the slow regulator solution which is identical to the singular linear quadratic solution [J10].

If instead of $B'QB > 0$ we have

$$B'(A')^iQA^iB = 0, \quad i = 0,1,\ldots,q-2,$$

$$B'(A')^{q-1}QA^{q-1}B \text{ is nonsingular,}$$

then, in successive cases, a singular perturbation analysis [J12] reveals increasingly impulsive behavior at the initial time $t = 0$, increased thickness of the boundary layer and lower dimensionality of the singular arc solution. It is often desired that the state trajectory $x_0(t,\epsilon)$ be bounded or at most step-like near $t = 0$ as $\epsilon$ tends to zero, for any initial state
A condition [J17] for bounded peaking is that rank $C(I-A)^{-1}B = \text{rank } CB$. Moreover, for a square invertible [J1] transfer-function matrix $G(\sigma) = C(\sigma I - A)^{-1}B$, perfect regulation (zero cost) is achieved as $\epsilon \to 0$, irrespective of the initial state, if and only if the transmission zeros [J7] of $G(\sigma)$ lie in the open left-half plane ($G(\sigma)$ is minimum phase) [J6,J14,J32]. Dual results on perfect state estimation in the presence of weak measurement noise are discussed in [J32,J45]. In the nonminimum phase case, there is a set of initial conditions for which perfect regulation cannot be achieved [J6,J52,J60]. This has to be viewed as an inherent performance limitation of nonminimum phase systems. The cheap control problem also forms the basis of a design procedure, using multivariable root loci [J5,J6], for the selection of the cost weighting matrices of the state regulator problem [J18,J36] and the output regulator problem [J41]. A partial cheap control problem, in which some but not necessarily all of the control inputs have arbitrarily small weighting in the performance index is studied in [J59]. Recent results [J61] establish that assumption (6.11) is not essential and the high-gain results extend to include all invertible systems [J1].

6.3. Variable Structure Systems

Variable structure systems [A3,A6,A7,J14] are systems (6.1) for which the so-called sliding mode on the switching surface $\zeta(x) = 0$ is achieved by the discontinuous feedback control, component-wise,
For \( \xi(x) = Cx = 0 \) and \( CB \) nonsingular the system (6.1) in sliding mode is governed by the "equivalent control"

\[
\begin{align*}
\dot{\xi} &= CAx + CBu = 0. \quad (6.14)
\end{align*}
\]

The feedback system (6.1), (6.13) is robust with respect to parameter variations, similar to high-gain systems, and the motion of (6.2) with (6.13) is identical to the slow motion of the high-gain feedback system (6.1), (6.2). The switching surfaces can be synthesized from the solution of either an eigenvalue assignment problem or a quadratic regulator problem for the slow subsystem [J14, J27, J28]. Developments of the variable structure approach include model-following systems [J30, J58], and servomechanism design with application to overspeed protection control [J57].
7. WEAK COUPLING AND TIME SCALES

The need for model simplification with a reduction (or distribution) of computational effort is particularly acute for large-scale systems involving hundreds or thousands of state variables, often at different geographical locations. Some form of decentralized modeling and control which exploits the weak interactions between subsystems is then required. While there are a number of approaches to the study of large-scale systems [A12], the success of any proposed decentralized scheme critically depends upon the choice of subsystems [M15]. In this respect, two-time-scale methods can be developed to aid the modeling process itself.

A fundamental relationship between time scales and weak coupling has been independently developed for power systems and Markov chains [K15, K18, L1, L2, L11, L13]. If the weak interactions of N "local" subsystems are treated as $O(\varepsilon)$, and if each subsystem has an equilibrium manifold (null space), then the local subsystems are decoupled in the fast time scale. However, they strongly interact in a slow time scale and form an aggregate model whose dimension is equal to the number of the local subsystems N. The system is thus decomposed into N+1 subsystems (N in the fast and one in the slow time scale).

7.1. Slow Coherency in Power Systems

The motivation to seek a system decomposition/aggregation along the above lines originates in the post-fault transient analysis of electromechanical models for multi-machine power systems [K6, K9, K15-K17] of the linearized form
\[ \dot{x} = M^{-1}(K_{\text{in}} + \epsilon K_{\text{ex}})x = A_{\epsilon}x \] (7.1)

where \( x \) represents the rotor angles, \( K_{\text{in}} \) are strong "internal," and \( \epsilon K_{\text{ex}} \) are weak "external connections." It has been observed that after a fault some machines have the tendency to swing together with an "in phase" slow motion. Such coherent machines are grouped into "coherent areas" which are then represented by "equivalent machines." Given the \( N \) slowest modes of \( A_{\epsilon} \), the machines \( i \) and \( j \) are said to be slowly coherent if the angles \( x_i - x_j \) contain only fast modes \([K_6, K_{15}]\). The practical problem in reduced order modeling of power systems is to find the groups of slow-coherent machines when the partition of \( A_{\epsilon} \) into \( K_{\text{in}} \) and \( \epsilon K_{\text{ex}} \) is not known, that is, the location of \( \epsilon \)-connections is not explicit. When the number of coherent areas is equal to \( N \), the number of slow modes of \( A_{\epsilon} \), the time-scale separation algorithm \([K6]\) for the "dichotomic" solution \( L = V_2 V_1^{-1} \) of the Riccati equation automatically groups the machines into areas. In the present notation, \( V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \) is an \( n \times N \) basis matrix for the \( N \) selected slow modes of \( A_{\epsilon} \). For this ideal, so-called \( N \)-decomposable case, the matrix \( L \) consists of rows which have only one nonzero element and it is equal to unity. This matrix \( L \) is termed a grouping matrix because it associates the remaining \( n-N \) angles \( x^2 \) to the \( N \) reference angles \( x^1 \). The angular differences \( x_i - x_j \) in the areas are fast variables defined by

\[ \eta(t) = x^2(t) - Lx^1(t) \] (7.2)

as a special case of (1.5). The slow variables are the area "centers of inertia"

\[ \xi_J = \sum_{i \in J} \frac{m_i}{m_j} x_i, \quad m_J = \sum_{i \in J} m_i, \quad J = 1, \ldots, N. \] (7.3)
In the new coordinates $\xi$, $\eta$ a near-decomposable power system appears in a standard singular perturbation form (1.12), (1.13). It is shown [K17] that for a 48-machine system, the area decomposition obtained for the linearized model remains valid in nonlinear simulations. Further details and applications of these methodologies are given in a recent monograph [K19].

7.2. Aggregation of Markov Chains

It has recently been established [L12] that the coherency method in power systems is analogous to a method for aggregation of Markov chains developed earlier in [L1,L2,L5,L7] and more recently in [L3,L4,L6,L10,L11,L13]. In Markov chains, high transition probabilities within a group of states, permit that for long-term studies this group be treated as an aggregate state. Much weaker transition probabilities to states of other groups have effect only after a long period of time, and can therefore be neglected over shorter periods, while for long-term studies an aggregate model can be used.

The probability vector $p$ of an $n$-state Markov chain with $N$ groups of strongly interacting states satisfies

$$\frac{dp}{dt} = p(A + \varepsilon B) \quad (7.4)$$

where $A = \text{diag}[A_1, A_2, \ldots, A_N]$. Weak interactions $\varepsilon B$ become significant after a long period in the fast time scale $\tau$. In the slow time scale $t = \varepsilon \tau$

$$\frac{dp}{dt} = p\left(\frac{1}{\varepsilon} A + B\right) \quad (7.5)$$
the fast transient is formed of separate transients within the strongly
coupled groups. After some time \( p_A \) decays to \( 0(\varepsilon) \), and from then on
\( p_B \) is no longer negligible with respect to \( \frac{1}{\varepsilon} p_A \). To make the slow and fast
parts of \( p(t) \) explicit, let

\[
y = pT, \quad z = p^s, \quad p = yV + zW
\]

(7.6)

where

\[
\begin{align*}
AT &= 0, & VA &= 0, & VT &= I_N \\
WT &= 0, & VS &= 0, & WS &= I_{n-N'}
\end{align*}
\]

(7.7)

Then (7.5) appears in the standard singular perturbation form

\[
\begin{align*}
\dot{y} &= yVBT + zWBT \\
\epsilon \dot{z} &= \epsilon yVBS + zW(A + \epsilon B)S
\end{align*}
\]

(7.8) (7.9)

where \( W(A + \epsilon B)S \) is a stable matrix. The slow subsystem is

\[
y_s = y_s(VBT - \epsilon VBS(WAS)^{-1}WBT) = y_s A_s
\]

(7.10)

For \( \epsilon = 0 \) it reduces to the aggregate chain proposed in [L1]. Its states
are the groups of the original states and \( y_j, J = 1, \ldots, N \) is the probability
for the original process to be in group \( J \). Based on the aggregate chain
(7.10), near-optimal policies have been obtained for controlled Markov
chain problems [L2,L4-L8,L10,L11,L13,L14]. The resulting optimization
algorithms are decentralized in the sense that fast subsystems compute
their controls "locally" with the aggregate coordinating necessary information
between subsystems and in a slow time scale. These results can be
considered as discretized versions of diffusion control processes [L3] and
are applicable to queueing problems [L9].
8. CONTROL OF LARGE SCALE SYSTEMS

Large scale systems are characterized by decentralization in available information, multiplicity of decision makers, and individuality of objective functions for each decision maker.

In recent years, singular perturbations have become an important tool for structuring large scale system models. Indeed the preceding section has shown that for a wide class of large scale systems, the notions of subsystems, their coupling, and time scales are interrelated and lead to a singularly perturbed model with a strongly-coupled slow "core" representing the system-wide behavior, and weakly-coupled fast subsystems representing the local behavior. The impact of this model on the design of control strategies is discussed in the section on multimodeling, while the last section deals with singularly perturbed differential games.

8.1. Multimodeling

Suppose that with original controls, or by their voluntary grouping, the linearized model of a large scale system exhibiting one slow and N fast subsystems can be written as

\[
\begin{align*}
\dot{x} &= A_{00}x + \sum_{i=1}^{N} A_{0i}z_i + \sum_{i=1}^{N} B_{0i}u_i \\
\epsilon_{i}\dot{z}_i &= A_{i0}x + A_{ii}z_i + \sum_{j=1}^{N} \epsilon_{ij}A_{ij}z_j + B_{ii}u_i
\end{align*}
\] (8.1)
where each fast subsystem is associated with a different singular perturbation parameter $\varepsilon_i$ and is weakly coupled to other fast subsystems through $\varepsilon_{ij}$.

In a situation like this, it is rational for a fast subsystem controller to neglect all other fast subsystems and to concentrate on its own subsystem, plus the interaction with others through the slow core. For the $i$-th controller, this is simply effected by setting all $\varepsilon$-parameters to zero, except for $\varepsilon_i$. The $i$-th controller's simplified model is then

\[
\dot{x}_i = A_i x_i + A_{0i} z_i + B_{0i} u_i + \sum_{j=1}^{N} B_{ij} u_j
\]

\[
\varepsilon_i \dot{z}_i = A_{10} x_i + A_{1i} z_i + B_{1i} u_i.
\]

Often (8.2) is all the $i$-th controller knows about the whole system. The $k$-th controller, on the other hand, has a different model of the same large scale system. This situation, called multimodeling, has been formulated and investigated in [M4,M10,M15] for deterministic problems, and in [M18,M19] for stochastic problems. Control $u_i$ can be divided into a slow part, which contributes to the control of the core, and a fast part controlling only its own fast subsystem. Pole-placement and Pareto solutions have been studied in [M4], while the closed-loop Nash solution has been analyzed in [M10].

In stochastic problems, a careful treatment is required to establish the validity of the multimodel problem. In addition to the usual difficulties encountered in modeling a fast stochastic variable, as discussed in Section 5, the problem is involved due to the presence of
nonclassical information patterns. To elucidate this aspect, consider the model

\[ \dot{x} = A_{00}x + \sum_{i=1}^{N} A_{0i}z_i + \sum_{i=1}^{N} B_{0i}u_i + G_0w \]

\[ \varepsilon_{i} \dot{z}_{i} = A_{i0}x + A_{ii}z_i + \sum_{j=1, j \neq i}^{N} \varepsilon_{ij} A_{ij} z_j + B_{ii}u_i + \sqrt{\varepsilon_{ii}} G_i w \]

where \( w(t) \) is white Gaussian noise. The information available \( y_i(t) \) and performance objective \( J_i \) of each decision maker are to be compatible with the multimodel assumption (8.2). This implies that

\[ y_{i0} = C_{i0}x + v_{i0} \]

\[ y_{ii} = C_{ii}z_i + \sqrt{\varepsilon_{ii}} v_{ii} \]

\[ y_i = [y_{i0} \ y_{ii}]' \quad ; \quad i = 1, 2, \ldots, N \]

\[ J_i = E \{ x_{i0}'T Q_{i0} x_{i0} + \sum_{i=1}^{N} z_{ii}'T Q_{ii} z_{ii} + \int_{0}^{T} (x'Q_{i0} x + z_{ii}'Q_{ii} z_{ii} + u_{i}'R_{i} u_i) dt \} \]

\[ i = 1, 2, \ldots, N \]

where \( v_{i0}(t) \) and \( v_{ii}(t) \) are white Gaussian noise processes, independent of each other and of the process noise \( w(t) \). Clearly, the problem defined by (8.3), (8.4) under any solution concept is one involving nonclassical information patterns for which no implementable finite-dimensional solution exists. Further assumptions need to be made about the information structure to obtain implementable solutions. In any case, the optimal strategies generally lead to the solution of a set of coupled integro-differential equations, and it is not clear whether their limiting solution would
correspond to the solution of the multimodel problem. To answer this question, a detailed investigation is undertaken in [M18, M19] for two important classes of quasi-classical information patterns. In [M18], well-posedness of the multimodel problem is established for a class of stochastic Nash games with a prespecified finite-dimensional compensator structure for each decision maker. The same fact is established in [M19] for static and dynamic team problems with sampled observations, under the one-step-delay observation sharing pattern. These results establish certain "robustness" of the multimodel problem to a class of solution concepts and information structures.

The assumption of weakly interacting fast subsystems is removed in [M6, M7, M8, M11, M17] where a sufficient "block D-stability" condition guarantees the asymptotic stability of the multiparameter boundary layers. Several tests for identifying classes of systems which satisfy this condition are reported in [M7, M8, M17]. The relationship of this condition with multiple time-scales is examined in [M7]. The main result of [M6] is a procedure for including limited, though not necessarily weak, interactions of fast subsystems. In [M11], applications to decentralized stabilization are examined and in [M16] a near-equilibrium solution to closed-loop Nash games is obtained. Extension of the "block D-stability" condition to a class of nonlinear systems which are linear in the fast variables, is obtained in [M14]. Related approaches to the multi time scale problem are reported in [M9] and [M12] for linear systems, and in [M20] for nonlinear systems. A series of papers [M1-M3, M5, M13] deal with the stability of large-scale singularly perturbed systems within a general framework, while [K8] uses singular perturbation techniques to study the hierarchical stability of power systems.
8.2. Singly Perturbed Differential Games

The references in this section report on the impact of singly perturbed reduced order models in the determination of simplified zero-sum, Nash and Stackelberg strategies. The main question investigated is one of well-posedness, whereby the limit of the performance using the exact strategies is compared to the limit of the performance using the simplified strategies. The simplified solution is said to be well-posed if the two limits are equal. Unlike the state feedback problem, the natural singular perturbation order reduction does not always lead to well-posed solutions of differential game problems. The answer crucially depends on the information structure of the game and its preservation in the reduced order problem since it is well-known that, in contrast to the optimal control problem, differential games admit different open- and closed-loop solutions.

The crucial role of information structure in reduced order modeling of differential games can be seen from the following two-player Nash game

\[ \begin{align*}
\dot{x} &= f(x,z,u_1,u_2,t); \quad x(t_o) = x_0 \\
\epsilon \dot{z} &= g(x,z,u_1,u_2,t); \quad z(t_o) = z_0 \\
\int_{t_o}^{t_f} L_i(x,z,u_1,u_2,t)dt; \quad i = 1,2. & (8.7)
\end{align*} \]

Necessary conditions for a closed-loop solution are

\[ \begin{align*}
\dot{p}_i &= -\nabla_x H_i - \left( \frac{\partial \psi_i}{\partial x}(t,x,z) \right) \nabla_{u_j} H_j; \quad p_i(t_f) = 0 \\
\epsilon \dot{q}_i &= -\nabla_z H_i - \left( \frac{\partial \psi_i}{\partial z}(t,x,z) \right) \nabla_{u_j} H_j; \quad q_i(t_f) = 0. & (8.8)
\end{align*} \]
and \( u_i = \psi_i(t,x,z) \) minimizes the Hamiltonian \( H_i(x,z,t,u_i,\psi_j) \), \( i,j = 1,2; i \neq j \).

For open-loop solution, (8.8) is replaced by

\[
\begin{align*}
\dot{p}_i &= -\nabla_x H_i; \quad p_i(t_f) = 0 \\
\dot{\psi}_i &= -\nabla_z H_i; \quad \psi_i(t_f) = 0.
\end{align*}
\]  

(8.9)

The partial derivative terms in (8.9) give rise to generally different open- and closed-loop solutions. Setting \( \epsilon = 0 \) reduces the necessary conditions for the closed-loop solution to

\[
\begin{align*}
0 &= g(x,z,t,u_1,u_2) \\
0 &= -\nabla_z H_i - \frac{\partial \psi_i}{\partial z}(t,x,z) \nabla_{u_j} H_j.
\end{align*}
\]  

(8.10)

Using (8.10) to eliminate \( z \) from (8.6), (8.7) implies that the optimal strategies \( \psi_1, \psi_2 \) are functions of \( t \) and \( x \) only. Thus \( \frac{\partial \psi_i}{\partial z} = 0 \). Suppose, however, that (8.6)-(8.8) are first solved and then \( \epsilon \to 0 \). If it is not true that

\[
\lim_{\epsilon \to 0} \nabla_z H_i = 0; \quad \forall t \in (t_o,t_f)
\]  

(8.11)

then the order reduction is generally not well posed. Under appropriate assumptions one may be able to show that

\[
\lim_{\epsilon \to 0} (\nabla_z H_i + \frac{\partial \psi_i}{\partial z} \nabla_{u_j} H_j) = 0; \quad \forall t \in (t_o,t_f).
\]  

(8.12)

The fact that (8.12) does not imply (8.11) is the reason for the ill-posedness of closed-loop Nash solutions [N2-N5]. In open-loop and partially
closed-loop (control functions of $x$ alone) Nash solutions, (8.12) does imply (8.11) because $\frac{\partial \psi_i}{\partial z} = 0$; and hence these solutions are well-posed [N5].

Another class of problems for which (8.12) implies (8.11) are LQ problems where the performance indices are modified (i.e., no penalty on the fast variable $z$) as in [N2], where also an alternative way of defining a reduced game whose closed-loop solution is the limit of the closed-loop solution of the full game is proposed. Related work on near-Nash feedback control of a composite system with a time-scale hierarchy is reported in [N11]. Analogous results for the Stackelberg problem are [N6,N9]. For zero-sum games, the normal order reduction is well-posed [N1], which is to be expected because zero-sum games have the same open- and closed-loop solutions.

Singular perturbation techniques are applied to pursuit-evasion problems in [N7,N8,N10,N12,N13].
CONCLUDING REMARKS

Research trends, discernible from this survey, can be briefly summarized in the broad categories of modeling, analysis, and design.

In modeling the trend is to go beyond the standard model discussed in Sections 1.1 and 1.2. The two more general classes are, first, models with known small parameters which do not multiply derivatives, and, second, models whose time-scale properties are caused by "hidden" parameters. Representatives of the first class are high-gain and some singular systems. In the second class are dynamic networks with coherent groups of states. Modeling issues in stochastic systems center around probabilistic scaling of fast phenomena. Scaling is also an issue in trajectory optimization problems.

In analysis, the trend is to encompass new classes of systems: discrete-time, stochastic and nonlinear; and to analyze their controllability, stability and other properties in separate time scales. In particular, singular perturbation method promises to be a powerful tool in robustness analysis of adaptive systems. In linear systems new frequency and geometric methods are being related to asymptotic methods.

Two-time-scale state feedback design methods are being extended to output feedback, observer and compensator design problems. The success of the composite control method for a class of nonlinear and stochastic systems motivates its extension to broader classes of systems. The multimodeling approach is one of the potential applications of time-scale methods to decentralized design of large scale systems and differential games.

These and related research directions depend on and will contribute to a deeper understanding of nonlinear and stochastic perturbation, bifurcation and averaging problems.
REFERENCES

A) Books and Surveys


B) Asymptotic Methods


C) Time-Scale Properties of Linear Systems


D) Discrete Time Systems


E) Linear State Feedback


F1 Output Feedback and Observers


Trajectory Optimization


1) Stochastic Systems. Filtering and Control


---

**High-Gain Feedback Systems**


---

**Time Scales in Networks & Power Systems**


L) Aggregation of Markov Chains


(M) Multimodeling and Stability of Large Scale Systems


Singularly Perturbed Differential Games


Singular Systems


**ADDITIONAL REFERENCES**

