ANALYTIC METHODS FOR THE APPROXIMATE SOLUTION OF SINGULARLY PER--Etc(U)
OCT 80  R E O'MALLEY

END
NUM
-81
DNC
An Introduction to Singular Perturbations

The books of Wasow (1965), Vasil'eva and Butuzov (1973), and O'Malley (1974) all discuss the asymptotic solution of singularly perturbed initial value problems for vector systems of the form

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

where \( \varepsilon \) is a small positive parameter and an asymptotic solution valid as \( \varepsilon \to 0 \) is sought. Assuming a little smoothness and assuming that the eigenvalues of the Jacobian matrix are strictly stable (i.e., have strictly negative real parts) everywhere implies that the asymptotic solution has the form

\[ x = \chi(t) + x_\varepsilon(t), \]

where \( \chi(t) \) satisfies the equations

\[ \begin{align*}
\dot{\chi} &= f(\chi, \varepsilon t, \varepsilon) \\
0 &= g(\chi, \varepsilon t, \varepsilon)
\end{align*} \]

prescribed on a finite or semi-infinite interval where \( t \geq 0 \). Here \( \varepsilon \) is a small positive parameter and \( x_\varepsilon(t) \) is an asymptotic solution valid as \( \varepsilon \to 0 \) is sought. Assuming a little smoothness and assuming that the eigenvalues of the Jacobian matrix are strictly stable (i.e., have strictly negative real parts) everywhere implies that the asymptotic solution has the form

\[ x = \chi(t) + x_\varepsilon(t), \]

where \( \chi(t) \) satisfies the equations

\[ \begin{align*}
\dot{\chi} &= f(\chi, \varepsilon t, \varepsilon) \\
0 &= g(\chi, \varepsilon t, \varepsilon)
\end{align*} \]

prescribed on a finite or semi-infinite interval where \( t \geq 0 \). Here \( \varepsilon \) is a small positive parameter and \( x_\varepsilon(t) \) is an asymptotic solution valid as \( \varepsilon \to 0 \) is sought. Assuming a little smoothness and assuming that the eigenvalues of the Jacobian matrix are strictly stable (i.e., have strictly negative real parts) everywhere implies that the asymptotic solution has the form

\[ x = \chi(t) + x_\varepsilon(t), \]

where \( \chi(t) \) satisfies the equations

\[ \begin{align*}
\dot{\chi} &= f(\chi, \varepsilon t, \varepsilon) \\
0 &= g(\chi, \varepsilon t, \varepsilon)
\end{align*} \]

prescribed on a finite or semi-infinite interval where \( t \geq 0 \). Here \( \varepsilon \) is a small positive parameter and \( x_\varepsilon(t) \) is an asymptotic solution valid as \( \varepsilon \to 0 \) is sought. Assuming a little smoothness and assuming that the eigenvalues of the Jacobian matrix are strictly stable (i.e., have strictly negative real parts) everywhere implies that the asymptotic solution has the form

\[ x = \chi(t) + x_\varepsilon(t), \]

where \( \chi(t) \) satisfies the equations

\[ \begin{align*}
\dot{\chi} &= f(\chi, \varepsilon t, \varepsilon) \\
0 &= g(\chi, \varepsilon t, \varepsilon)
\end{align*} \]

prescribed
and $\eta(t)$ is an asymptotically stable solution of the (parameterless) "boundary layer" problem

\[
\begin{align*}
\frac{dn}{dt} &= g(M(0), N(0) + \eta, 0, 0), \quad t \geq 0 \\
\eta(0) &= \beta(0) - N(0).
\end{align*}
\]

(4)

The $O(\epsilon)$ terms represent small residuals bounded in magnitude (on fixed finite $t$ intervals) by $C\epsilon$, for some bounded constant $C$, provided $\epsilon$ is sufficiently small. Hoppensteadt (1966) further showed that such results continue to hold for all $t \geq 0$ provided $M$ and $N$ decay exponentially as $t \to \infty$.

It is essential to realize the substantial order reduction achieved through the approximations (2). The implicit function theorem guarantees a locally unique solution

(5)

\[ N(t) = \phi(M(t), t) \]

of the nonlinear system $g(M, N, t, 0) = 0$. (In practice, numerical or symbolic solution methods would be used to obtain $\phi$, especially when $\beta$ has many components and the order reduction is substantial.) Knowing $\phi$, $M(t)$ follows from solving the initial value problem

(6)

\[ \dot{M} = F(M, t) \equiv f(M, \phi(M, t), t, 0), \quad M(0) = \alpha(0). \]

With our hypotheses (or, e.g., the weaker ones of Howes (1979)), a unique solution $\eta(t)$ of (4) will decay to zero as the "stretched variable" $t/\epsilon \to \infty$. Thus, the $\eta$-term allows nonuniform convergence (as $\epsilon \to 0$) of the $\beta$ variable in a narrow "boundary layer" region near $t = 0$. Since $\eta$ becomes negligible as $\epsilon \to 0$ for each fixed $t > 0$, the limiting solution away from $t = 0$ is uniquely determined by the solution $M(t)$ of the reduced problem (6) and by $N(t) = \phi(M(t), t)$. 
The reduction of order is completely analogous to the pseudo-steady state hypothesis of chemical kinetics. In the celebrated enzyme kinetics theory of Michaelis and Menton, an \( \epsilon \) can be introduced (as in (1)) as the ratio of the initial enzyme to the initial substrate concentrations (cf. Heinekin et al. (1967) and Murray (1977)). Solving the reduced problem, instead of the original problem with its fast initial transient (represented by \( n(t/\epsilon) \) in (2)) is most advantageous for purposes of numerical integration (cf. Aiken and Lapidus (1974) and Miranker (1980)). The full system will be stiff for small values of \( \epsilon \), and stability requirements for Euler's method and similar explicit schemes will force one to use a small stepsize for \( t > 0 \), even though the solution is smooth there. As Shampine (1980) and Dahlquist et al. (1980) point out, implicit schemes for the full system may not so restrict the stepsize and may actually be solving systems closely linked to the reduced problem (3).

The preceding hypotheses are restrictive, and it should be realized that related (though less complete) results are available in other (more complicated) contexts (cf., e.g., Bavinck and Grasman (1969), Hoppensteadt and Miranker (1976), O'Malley and Flaherty (1980), Bobisud and Christenson (1980), and O'Malley (1980)). If, for example, \( g_y \) is (everywhere) singular, we cannot expect to uniquely determine the limiting solution from the reduced problem alone. If \( g_y \) has purely imaginary eigenvalues, we can expect rapidly oscillating solutions. If \( g_y \) has eigenvalues with both positive and negative real parts, the boundary layer system is conditionally stable and we can expect nonuniform convergence (boundary layer behavior) for corresponding two-point problems at both endpoints. Considerable complexity can be expected when (where) the eigenvalue structure of \( g_y \) changes (both bifurcation theory and turning point theory may be called for).

Real problems naturally simultaneously involve many small parameters.
Detailed theories are available for problems with several interrelated parameters (cf. O'Malley (1969)) and, even, for t-dependent parameters (cf. Gingold (1980)). The one parameter \( \epsilon \) in (1) should be interpreted as an aggregate parameter used to distinguish the "fast" variables \( \beta \) from the "slow" variables \( \alpha \). More refined models could use several small parameters to distinguish different groupings of fast variables. Indeed, a realistic modeler often introduces a hierarchy of models, increasing in dimensionality, allowing increasingly accurate models of the same system. In practice, researchers will make different decisions regarding both the overall number of components employed and the appropriate number of variables for reduced-order models.

The structure of the singular perturbation problem (1) is very special, because the components \( \alpha \) are clearly separated (via the parameter \( \epsilon \)) from the \( \beta \) components which are anticipated to be more rapidly varying, at least in narrow boundary or interior transition regions. In applied problems, practitioners often know which variables are "fast." There is, for example, a natural dichotomy when one deals with dynamics involving both electrical and, relatively sluggish, mechanical components. Analogous obvious splittings commonly occur in chemical contexts as well. Often, however, variables are not completely decoupled into fast and slow sets and a preliminary decoupling procedure must be used before singular perturbation concepts become helpful (i.e., before a meaningful small parameter \( \epsilon \) can be introduced.) Numerical decoupling methods are currently being developed by Enright and Kamel and by Dahlquist and Söderlind (preliminary work is reported in Enright and Kamel (1979) and Söderlind (1979)). We also note the emphasis of Dahlquist et al. (1980) on scaling techniques for nonlinear differential systems in dimensionless variables (cf. Lin and Segel (1974) as well). Our experience relates primarily to models for aircraft engines, structural dynamics, and power
distribution and generation (cf. O'Malley and Anderson (1980), Anderson and Hallauer (1980), Kokotovic et al. (1980), and Girijashankar et al. (1980)). We expect, however, that the underlying strategy will be applicable in chemical systems and a wide spectrum of other fields.

**Linear Systems without Explicit Parameters**

We shall consider large-scale linear systems of the form

\[ \dot{x} = A(t)x + u(t) \]

where \( u(t) \) is a forcing function, not a feedback control. Such a system could, of course, be obtained from linearizing a nonlinear system about a nominal trajectory. The initial behavior of any solution will depend on the eigenvalues of \( A(0) \). Let us suppose that

\[(H1). \quad n_1 \text{ of the eigenvalues of } A(0) \text{ are small in magnitude compared to the remaining } n_2 = n - n_1 \text{ eigenvalues.}\]

(A finer subdivision may be appropriate for a more careful model.) If such a slow/fast split in solution modes for the homogeneous system could be maintained, we might hope (as for the singularly perturbed system (1)) that the approximate dynamics of the full \( n \)th order system (7) (away from transient regions) could be given by a lower \( n_1 \)th order system.

We will first try to transform system (7) to a decoupled form by use of a transformation

\[ y = T(t)x = T_2(t)T_1(t)x = T_2(t)z \]

where \( T_1 \) and \( T_2 \) are the block-triangular matrices.
\[ T_1(t) = \begin{bmatrix} I_{n_1} & 0 \\ L(t) & I_{n_2} \end{bmatrix} \]

and

\[ T_2(t) = \begin{bmatrix} I_{n_1} & K(t) \\ 0 & I_{n_2} \end{bmatrix} \]

specified by the decoupling matrices L(t) and K(t). Since these matrices are easily inverted, we immediately have the inverse

\[ T^{-1} = \begin{bmatrix} I_{n_1} & -K \\ -L & I_{n_2} + LK \end{bmatrix} \]

If we subdivide \( A = (A_{ij}) \) after its first \( n_1 \) rows and columns, the intermediate variable \( z = T_1x \) will satisfy

\[ \dot{z} = B(t)z + T_1(t)u \]

with the compatibly partitioned \( B = (B_{ij}) \) being upper block triangular provided \( L(t) \) satisfies the matrix Riccati equation

\[ \dot{L} = A_{22} L - LA_{11} + LA_{12} L - A_{21} \]
Specifically, we note that

\[
B_{11} = A_{11} - A_{12} L, \quad B_{12} = A_{12}, \quad B_{21} = 0, \quad \text{and} \quad B_{22} = A_{22} + LA_{12}.
\]

We also note that transformations such as (9) occur in many classical contexts, including reduction-of-order (cf. Hartman (1964)) and asymptotic theory (cf. Wasow (1965)). Here, the matrix \(L\) is generally nonsquare and the Riccati equation has time-varying coefficients, in contrast to the symmetric and time-invariant case most familiar in control and estimation theory (cf. Van Dooren (1980)). In the general context needed here, conditions guaranteeing existence of the Riccati solution do not seem to be known. Most available theory is contained in Reid (1972), Medanic (1979), and Bart et al. (1979). Blowup of solutions can, of course, be detected in numerical solutions and "integration to blowup" is actually used to calculate eigenlengths (cf. Nelson and Elder (1977)). It would be helpful to know which solution of the Riccati equation one should use (i.e., which initial value to impose for \(L(0)\)) to prolong the interval of existence or to optimize the stability of a numerical solution technique.

Using the second transforming matrix \(T_2(t)\) and splitting the variable \(y\) after its first \(n_1\) rows provides the decoupled system

\[
\begin{pmatrix}
y_1 \\
y_2
\end{pmatrix}
= \begin{pmatrix}
B_{11} & 0 \\
0 & B_{22}
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_2
\end{pmatrix}
+ T(t)u
\tag{13}
\]

provided the matrix \(K\) satisfies the linear system

\[
K = B_{11} K - KB_{22} - A_{12}.
\tag{14}
\]
A block diagonalization is actually unnecessary to decouple modes of the unforced system since $T_1$ already achieved a block triangularization, but we proceed to diagonalize because obtaining $K$ is relatively easy.)

Altogether, then, we have shown that solving the original vector or matrix system (7) (thereby finding a fundamental matrix for the homogeneous problem) is equivalent to solving the four systems:

\[
\begin{align*}
\dot{L} &= A_{22}L - LA_{11} + LA_{12}L - A_{21} \\
\dot{K} &= (A_{11} - A_{12}L)K - K(A_{22} + LA_{12}) - A_{12} \\
y_1 &= (A_{11} - A_{12}L)y_1 + (I_{n_1} + KL)u \\
y_2 &= (A_{22} + LA_{12})y_2 + (L I_{n_2})u
\end{align*}
\]

(16)

We hope to convince the reader, that with the right hypotheses, we can determine solutions to these systems easier than solving the original system directly.

Note that if $L(0) = 0$, $T_1(0)$ will be a similarity transformation, and $B(0)$ and $C(0)$ will have the same eigenvalues as $A(0)$. We would have succeeded in separating the initially fast (and slow) modes of the transformed homogeneous system if we could guarantee that the $n_1$ smallest eigenvalues of $B(0)$ (identified by hypothesis (H1)) are those of $B_{11}(0)$ (the $n_2$ largest eigenvalues will then automatically be those of $B_{22}(0)$). The algebraic Riccati equation $\dot{L}(0) = 0$ for $L(0)$ will have many solutions, but that solution $L(0)$ which makes the eigenvalues of $B_{11}(0) = A_{11}(0) - A_{12}(0) L(0)$ coincide with the $n_1$ small eigenvalues of $A(0)$ is unique. This result is well-known for the symmetric algebraic Riccati equation (cf. Martensson...
(1971)) and it was proved in Anderson (1979) in our nonsquare context. Anderson also showed that the appropriate \( L(0) \) could be readily obtained by iterating. For computational purposes, one defines

\[
\varepsilon_{i+1} = (A_{22}(0) + \varepsilon_i A_{12}(0))^{-1} (\varepsilon_i A_{11}(0) + A_{12}(0))
\]

for each \( i \geq 0 \) and lets \( L(0) = \lim_{i \to \infty} \varepsilon_i \). This scheme turns out to be robust with respect to initial iterates \( \varepsilon_0 \) and it converges like a geometric series with ratio \( \mu \) proportional to the quotient of the largest small to the smallest large eigenvalue of \( A(0) \). An alternative representation is \( L(0) = -M_2 M_1^{-1} \) where \( M_1 \) spans the \( n_1 \) dimensional (generalized) eigenspace of the small eigenvalues of \( A(0) \). (A less convenient computational approach to finding \( L(0) \) would then be to approximate a basis for this eigenspace and form the indicated quotient.) This formula makes it clear that a rearrangement of the components of \( x \) may be necessary to assure the invertibility of \( M_1 \). Any ill-conditioning of \( M_1 \) can also cause difficulties (cf. Watkins (1980)).

We shall assume that our system is "two-time scale" after the transformation, viz

\((H2)\) the solution \( L(t) \) of the initial value problem for the matrix Riccati equation remains bounded throughout a fixed interval \( 0 \leq t \leq T \) and the eigenvalues of the matrix \( B_{11}(t) = A_{11} - A_{12} L \) remain small in magnitude compared to those of \( B_{22}(t) = A_{22} + L A_{12} \) throughout the interval.
Under this hypothesis, the free (i.e., unforced) response of the first $n_1$ components of the decoupled system for $y$ will be much slower than that of the last $n_2$ components. (For two point problems, $T$ will represent the right endpoint, while it should be a typical length scale for initial value problems.) In practice, it may be advisable to reinitialize $L(t)$ from time-to-time, just as the Scott and Watts' (1977) boundary value code reorthogonalizes to maintain stability.

Specializing to initial-value problems for (7), let us also impose a "fast-mode stability" hypothesis, viz

$$(H3) \quad \text{the eigenvalues of } B_{22}(t) \text{ all have large strictly negative real parts throughout } 0 \leq t \leq T.$$ 

In using this assumption, we are close to succumbing to a popular heresy: that eigenvalue stability implies stability. This is incorrect for time-varying systems, as Hoppensteadt (1966) points out by counterexample. As our initial statement shows, however, it is true for singularly perturbed systems with appropriately smooth coefficients. Kreiss (1978, 1979) presents a counterexample, but not in the special structure generally assumed. We could carry out an explicit and rigorous analysis involving the two small parameters implicitly introduced by assumptions (H2) and (H3), but this would require definite assumptions about parameter dependence that we wish to avoid being explicit about.

We note that the differential system (14) for $K$ is opposite in stability to the linearized equation

$$(18) \quad \dot{\varepsilon} = -B_{11} \varepsilon + \varepsilon B_{22}$$
for L. Since (H2) implies that the initial value problem for L(t) is well-behaved, (backward) numerical integration of a terminal value problem for K is suggested. Taking \( K(T) = 0 \) uniquely determines \( K(T) \) since the resulting system for \( K(T) \) is a linear one whose coefficient matrices \( B_{11}(T) \) and \( B_{22}(T) \) have no eigenvalues in common (cf., e.g., Bellman (1970)).

Because (H-3) implies that \( B_{22}(t) \) has eigenvalues with large negative real parts, the system (14) for \( K(t) \) can be rewritten in the traditional singular perturbation form with a small-parameter multiplying the derivative. Indeed, the usual theory implies that the limiting solution for \( t < T \) will satisfy a reduced system \( \dot{K}(t) = 0 \) (i.e. a pseudo-steady state approximation will apply.). Because we have \( \dot{K}(T) = 0 \), we'd expect no discontinuity at \( T \). \( K(t) \) can, then, be readily calculated as the limit of the iteration

\[
K_{j+1} = -B_{22}^{-1} (A_{12} - B_{11} K_j)
\]

(The derivative term can also be retained as a small perturbation in the right-hand-side.) Thus, the differential system (14) for \( K(t) \) has been essentially replaced by an easier-to-solve linear algebraic equation \( \dot{K}(t) = 0 \). In conducting numerical experiments, we've found that the pseudo-steady state algebraic Riccati equation \( \dot{L}(t) = 0 \) is quite a good approximation for finding \( L(t) \) for \( t > 0 \), but we are not willing to generally recommend that drastic simplification. We note that the linearization (18) (about a solution \( L \) of our Riccati system) is singularly perturbed under our hypotheses, but the original system (11) is not of singular perturbation form. Söderlind and Dahlquist (personal conversation) are, however, experiencing success in using piecewise constant solutions to algebraic Riccati equations in similar contexts.
Once \( L \) and \( K \) are (approximately) determined, there remain linear systems for \( y_1 \) and \( y_2 \). The "fast" system

\[
y_2 = B_{22}(t)y_2 + [L_{1n_2}] u(t)
\]

will be amenable to a pseudo-steady state approximation \( y_2 = 0 \) away from \( t = 0 \) provided the resulting approximate \( y_{2s} \) is slowly-varying. Thus, we'll finally ask

\[
\text{(H4) The approximation } y_{2s}(t) = -B_{22}^{-1} (L_{1n_2}) u(t) \text{ is slowly-varying}
\]

(i.e. \( y_{2s} \) is small) throughout \( 0 \leq t \leq T \). Even though \( B_{22} \) is large, note that this might be difficult to achieve if \( u(t) \) were rapidly varying. If the initial behavior of \( y_2 \) is important, one needs to add a "fast" initial layer corrector \( y_{2f} \) to the "slow-mode" solution \( y_{2s} \). The corrector will be a decaying solution to \( y_{2f} = B_{22}(t)y_{2f} \). Because of our hypothesis (H3), this system would only need to be integrated - with a small stepsize - over a short initial interval. This correction will be initially important because \( y_{2s}(0) \neq y_{2}(0) \) will generally force an initial nonuniform convergence in \( y_2 \) components.

The remaining "slow" system (16) for \( y_1 \) will be integrated as an initial value problem using the transformed initial condition

\[ y_1(0) = [I_{n_1} + K(0)L(0)K(0)] x(0). \]

Having found (or approximated) \( L, K, y_1, \) and \( y_2 \), the inverse transformation

\[ x(t) = \begin{bmatrix} I_{n_1} & -K(t) \\ -L(t) & I_{n_2} + L(t)K(t) \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} \]
provides (an approximation for) the original variable $x(t)$. Altogether, for $t > 0$, our approximations allow us to obtain an approximate fundamental matrix for $x$ by integrating the systems for $L$ and $y_1$ and using algebraic systems for $K$ and $y_2$.

In practice, we've solved initial value problems for vectors of dimensions four to sixteen (cf. O'Malley and Anderson (1980)) with two to five corresponding slow modes providing order reduction. Away from the initial point, good approximations were soon achieved by the low order models even though the parameters measuring time-scale separation and fast-mode stability were only moderately small. For such low order models, we have the advantage of being able to compare results to accurate solutions obtained numerically with small stepsizes. More testing is needed for higher dimensional models, where it is essential to use one's physical insight to a maximum to make initial decisions on which components are expected to be predominantly slow or fast. More effort should also be spent on boundary value problems. There the advantages of our asymptotic techniques will be even greater than for initial value problems.

Use of some of these concepts for nonlinear problems has been reported by Kreiss (1979), Girijashankar et al. (1980), Kokotovic et al. (1979), and Watkins (1980).

Acknowledgements: This work has benefitted from the ongoing collaboration with Professor Leonard Anderson of Virginia Polytechnic Institute. It was supported in part by the Office of Naval Research under Contract No. N00014-76-C-0326.
References


**Report Title:** Analytical Methods for the Approximate Solution of Singularly Perturbed Systems

**Author:** R.E. O'Malley, Jr.

**Performing Organization:** Program in Applied Mathematics, Mathematics Bldg., University of Arizona, Tucson, Arizona 85721

**Controlling Office:** Mathematics Branch, Office of Naval Research, Arlington, Virginia

**Report Date:** October 1980

**Number of Pages:** 16

**Security Classification:** Unclassified

**Distribution Statement:** Approved for public release. Distribution unlimited.

**Key Words:** Singular perturbations, reduced-order modeling, model decoupling, Michaelis-Menten kinetics.

**Abstract:** This presentation surveys the singular perturbation technique for initial value problems for large systems of ordinary differential equations. A decoupling transformation to separate fast from slow modes of unforced systems is introduced, and approximations based on the singular perturbation methodology are developed.