Engineering Notes

ENGINEERING NOTES are short manuscripts describing new developments or important results of a preliminary nature. These Notes cannot exceed 6 manuscript pages and 3 figures; a page of text may be substituted for a figure and vice versa. After informal review by the editors, they may be published within a few months of the date of receipt. Style requirements are the same as for regular contributions (see inside back cover).

Computational Singular Perturbation Method for Dynamic Systems

Mark D. Ardema*
Santa Clara University, Santa Clara, California 95053

Introduction

CONVENTIONAL singular perturbation analysis of dynamic systems relies ideally on identifying small (or large) parameters in the system equations. For reviews of singular perturbation methods, see for example Refs. 1-3. In many important problems, however, there are no physically identifiable small parameters and the system must be structured in terms of artificially imposed time scales prior to analysis. This has several disadvantages. First, the time-scaling procedure is rather ad hoc and relies on physical intuition and system insight. Second, the time-scale properties of a system are dependent on initial conditions and other problem data. For a discussion of these issues as they relate to aircraft trajectory optimization, see Ref. 4.

Recently, a new method of singular perturbation analysis, called computational singular perturbation (CSP) by its authors, has been developed to study complex chemical reaction systems.[5,6] This method does not rely on any a priori knowledge of system behavior. Rather, it produces time-scale information and corresponding approximate system behavior in the course of a numerical solution. It provides a family of efficient integration algorithms for stiff differential equations by always choosing the largest appropriate time step. Finally, it automatically responds to changes in initial conditions and other problem data.

In this Note, we will concisely review the CSP method and illustrate its application to general dynamic systems by applying the method to two simple examples. The terminology and symbolism of Refs. 5 and 6 will be used as much as possible.

Transformation of Dynamic Systems

We consider autonomous dynamic systems of the form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = g(y), \qquad y(0) = y_0, \qquad y \epsilon R^N \tag{1}$$

In Refs. 5 and 6, $g(y)$, called the global reaction rate vector, is the sum of several, say R, chemical reactions:

$$g(y) = \sum_{r=1}^{R} S_r F^r \tag{2}$$

Following Lam and Goussis,[5,6] $S_i F^i$ is called reaction group i; S_i is the stoichiometric vector, and F^i the reaction rate for reaction i, respectively. An observation fundamental to the CSP method is that the representation of $g(y)$ given by Eq. (2), while perhaps physically meaningful, is not unique. Indeed, a vector function may always be written in the form

$$g(y) = \sum_{i=1}^{N} a_i f^i \tag{3}$$

where the vectors a_i are a set of N linearly independent basis vectors and the f^i are the projections of g along the a_i.

Given a representation of g such as Eq. (2), we now derive a formula for the reaction rates f^i for any set of basis vectors a_i. Let b^i be the set of N vectors orthonormal to a_j:

$$b^i \cdot a_j = \delta^i_j, \qquad i,j = 1,2,\ldots,N \tag{4}$$

where δ^i_j is the Kronecher delta. Now take the dot product of b^i and g and use Eqs. (3) and (4):

$$b^i \cdot g = b^i \cdot \sum_{j=1}^{N} a_j f^i = \sum_{j=1}^{N} (b^i \cdot a_j) f^i = \sum_{j=1}^{N} \delta^i_j \cdot f^j = f^i \qquad i = 1,2,\ldots,N \tag{5}$$

Consequently, using Eq. (2),

$$f^i = b^i \cdot g = b^i \cdot \sum_{r=1}^{R} S_r F^r = \sum_{r=1}^{R} (b^i \cdot S_r) F^r \qquad i = 1,2,\ldots,N \tag{6}$$

which is the desired relation.

To illustrate these concepts, we introduce two simple examples. We will return to these examples later and apply the CSP method.

Example 1 is a linear oscillator with small mass. Taking unit values for the spring and damping constants and denoting the mass by ε, the dynamic system is

$$\frac{\mathrm{d}y_1}{\mathrm{d}t} = g_1 = -\frac{1}{\varepsilon} y_1 - \frac{1}{\varepsilon} y_2 + \frac{1}{\varepsilon}\sigma \tag{7a}$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = g_2 = y_1 \tag{7b}$$

where y_1 is the mass velocity, y_2 the displacement, and $\sigma(y_1, y_2)$ the applied force. One possible reaction group representation of Eq. (7) has stoichiometric vectors and reaction rates as follows:

$$S_1 = \begin{pmatrix} -1/\varepsilon \\ 1 \end{pmatrix}, \qquad S_2 = \begin{pmatrix} -1/\varepsilon \\ 0 \end{pmatrix}, \qquad S_3 = \begin{pmatrix} 1/\varepsilon \\ 0 \end{pmatrix}$$

$$F^1 = y_1, \qquad F^2 = y_2, \qquad F^3 = \sigma$$

Noting that S_1 and S_2 are linearly independent, for an alternative representation we may pick

$$a_1 = \begin{pmatrix} -1/\varepsilon \\ 1 \end{pmatrix}, \qquad a_2 = \begin{pmatrix} -1/\varepsilon \\ 0 \end{pmatrix}$$

Received Oct. 23, 1989; revision received Dec. 7, 1989.

*Professor and Chairman, Department of Mechanical Engineering. Associate Fellow AIAA.

From Eq. (4),

$$b^1 = (0 \quad 1), \qquad b^2 = (-\varepsilon \quad -1)$$

and then from Eq. (6)

$$f^1 = y_1, \qquad f^2 = y_2 - \sigma$$

Equation (3) gives the global reaction rate vector in terms of the new reaction groups as

$$g = \begin{pmatrix} (-1/\varepsilon)(y_1) + (-1/\varepsilon)(y_2 - \sigma) \\ (1)(y_1) + (0)(y_2 - \sigma) \end{pmatrix}$$

which is clearly equivalent to the original formulation.

Now consider the free oscillator

$$\frac{\mathrm{d}y_1}{\mathrm{d}t} = g_1 = -\frac{1}{\varepsilon} y_1 - \frac{1}{\varepsilon} y_2 \tag{8a}$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = g_2 = y_1 \tag{8b}$$

Analysis by a conventional singular perturbation technique, such as matched asymptotic expansions, reveals that this problem has two time scales. On the first time scale, $t = \mathcal{O}(\varepsilon)$, an approximate set of equations (valid to zero order) is

$$\frac{\mathrm{d}y_1}{\mathrm{d}\tau} = -y_1 - y_2 \tag{9a}$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}\tau} = 0 \tag{9b}$$

where $\tau = t/\varepsilon$. These equations are termed the boundary layer, or zero-order inner equations. On the other time scale, $t = \mathcal{O}(1)$, approximate relations are given by the reduced, or zero-order outer equations:

$$0 = y_1 - y_2 \tag{10a}$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = y_1 \tag{10b}$$

Now suppose that the singularly perturbed nature of Eq. (8) is not known a priori. Is it still possible, with a numerical algorithm, not only to efficiently solve the equations but to deduce the time-scale properties? It is clear that the $\mathcal{O}(1)$ terms in Eq. (8) are the cause of the system behavior in the outer region; in fact simply neglecting them gives the boundary-layer Eq. (9). It is also clear that the $\mathcal{O}(1/\varepsilon)$ terms are the cause of the boundary-layer behavior; neglecting them gives, however,

$$\frac{\mathrm{d}y_1}{\mathrm{d}t} = 0$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = y_1$$

which are not the same as the reduced Eq. (10). In the CSP method, the global reaction rate vector is transformed into a sum of ordered reaction groups such that neglecting one of the reaction groups gives the correct second-order differential equation form of the reduced system, namely:

$$\frac{\mathrm{d}y_1}{\mathrm{d}t} = y_2 \tag{11a}$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = -y_2 \tag{11b}$$

Example 2 is a nonlinear system considered first by Wasow[7]:

$$\frac{\mathrm{d}y_1}{\mathrm{d}t} = \frac{y_2^2 - y_1^2}{\varepsilon}, \qquad y_1(0) = 0 \tag{12a}$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = y_1, \qquad y_2(0) = 1 \tag{12b}$$

This system was analyzed and solved by Wasow using Vasileva's method,[8] and by Ardema[1] using the method of matched asymptotic expansions. The initial conditions stated in Eq. (12) have the implication that $y_1(t) \geq 0$ and $y_2(t) > 0$ for all $t \geq 0$. For this system, the boundary-layer equations are

$$\frac{\mathrm{d}y_1}{\mathrm{d}\tau} = y_2^2 - y_1^2 \tag{13a}$$

$$\frac{\mathrm{d}y_1}{\mathrm{d}\tau} = 0 \tag{13b}$$

and the reduced equations are

$$0 = y_2^2 - y_1^2 \tag{14a}$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = y_1 \tag{14b}$$

As before, simply discarding the boundary-layer terms in Eq. (12) gives

$$\frac{\mathrm{d}y_1}{\mathrm{d}t} = 0$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = y_1$$

whereas the correct second-order form of the reduced equations is

$$\frac{\mathrm{d}y_1}{\mathrm{d}t} = y_2 \tag{15a}$$

$$\frac{\mathrm{d}y_2}{\mathrm{d}t} = y_1 \tag{15b}$$

Computational Singular Perturbation Algorithm

In the CSP method of Lam and Goussis, $g(y)$ is represented "optimally," that is, as a sum of reaction groups ordered according to the rate of change of their reaction rates. The fastest reaction rates are monitored according to their relative numerical contribution to g; when this contribution is insignificant, the corresponding reaction group is simply dropped from the global reaction vector, and the integration time step is increased to a value consistent with the next fastest reaction group. The process then continues as before with the remnant of g.

We begin by writing g as the sum of a fast and a slow part:

$$g = g_f + g_s \tag{16}$$

and then put g_f in terms of fast reaction groups

$$g_f = \sum_{i=1}^{n} a_i f^i \tag{17}$$

where n is the number of fast reactions under consideration for elimination. As before, introduce a set of vectors b^i such that

a_i and b^i are dual sets of real basis vectors of the fast subspace of g:

$$b^i \cdot a_j = \delta^i_j, \qquad i,j = 1,2,\ldots,n \tag{18}$$

and require that the g_s be orthogonal to b^i:

$$b^i \cdot g_s = 0, \qquad i = 1,2,\ldots,n$$

Taking the dot product of b^i and g and using Eqs. (2) and (16–18) gives an expression for f^i:

$$b^i \cdot \sum_{r=1}^{R} S_r F_r = b^i \cdot \sum_{j=1}^{n} a_j \cdot f^j + b^i \cdot g_s$$

$$f^i = b^i \cdot g = \sum_{r=1}^{R} (b^i \cdot S_r) F^r \tag{19}$$

The slow part of the global reaction rate vector is determined from Eqs. (16), (17), and (19) as follows:

$$g_s = g - \sum_{i=1}^{N} a_i \cdot (b^i \cdot g) = \left(I - \sum_{i=1}^{n} a_i b^i \right) g \tag{20}$$

where I is the identity matrix.

It is the percentage rate of change of the f^i that is needed to order the reaction groups; to this end, differentiate Eq. (19) to give

$$\frac{\mathrm{d}f^i}{\mathrm{d}t} = \frac{\mathrm{d}b^i}{\mathrm{d}t} \cdot g + b^i J \cdot g, \qquad i = 1,2,\ldots,n \tag{21}$$

where

$$J = \frac{\partial g}{\partial y} = \text{Jacobian of } g \tag{22}$$

Now define a matrix $\Lambda = \{\Lambda^i_j\}$ by

$$\frac{\mathrm{d}f^i}{\mathrm{d}t} = \sum_{j=1}^{n} \Lambda^i_j f^j, \qquad i = 1,2,\ldots,n \tag{23}$$

We call Λ the rate matrix, because if it is diagonal, the diagonal elements are the normalized rates of change of the reaction rates f^i. Then, from Eqs. (19), (21), and (23),

$$\sum_{j=1}^{n} \Lambda^i_j b^j \cdot g = \frac{\mathrm{d}b^i}{\mathrm{d}t} \cdot g + b^i J \cdot g$$

which implies that

$$\frac{\mathrm{d}b^i}{\mathrm{d}t} = \sum_{j=1}^{n} \Lambda^i_j b^j - b^i J, \qquad i = 1,2,\ldots,n \tag{24}$$

Taking the dot product of this equation with a_k and using Eq. (18) gives an expression for Λ^i_k:

$$\Lambda^i_k = b^i J \cdot a_k + \frac{\mathrm{d}b^i}{\mathrm{d}t} \cdot a_k, \qquad i,k = 1,2,\ldots,n \tag{25}$$

Now note from Eq. (25) that if the a_k are constant vectors, then the b^i are as well and Λ^i_k becomes simply

$$\Lambda^i_k = b^i J \cdot a_k, \qquad i,k = 1,2,\ldots,n \tag{26}$$

and that if, additionally, b^i is a left eigenvector of J with eigenvalue μ_i, then

$$\Lambda^i_k = \mu_i \delta^i_k \tag{27}$$

or

$$\Lambda = \begin{bmatrix} \mu_i & & & 0 \\ & \mu_2 & & \\ & & \ddots & \\ 0 & & & \mu_n \end{bmatrix}$$

Consequently, the eigenvalues of the Jacobian J are directly related to, but generally not identical with, the "optimal" normalized reaction rates. This is a familiar result from singular perturbation theory for linear systems. To summarize, given n and $b^i(0)$, Eq. (24) may be integrated for $b^i(t)$ and then $a_i(t)$ determined from Eq. (18).

When discretizing this procedure for use in a numerical algorithm, there are several options. One algorithm, described by Lam and Goussis,[6] is generally as follows:

1) At $t = 0$, choose $n = 1$, $a_1 = \alpha_i(0)$, and $b^1(0) = \beta^1(0)$ where

$$\beta^1(0)\, J(0) = \mu_1(0) \beta^1(0) \tag{28a}$$

$$J(0) \alpha_1(0) = \mu_1(0) \alpha_1(0) \tag{28b}$$

and $\mu_1(0)$ is the largest eigenvalue in magnitude of $J(0)$, assumed to be negative. (Boundary-layer stability is of course still required. In Lam and Goussis's terminology, a negative eigenvalue signals a relaxation group and potential boundary-layer behavior, whereas a positive eigenvalue signals an explosive group.)

2) Integrate Eqs. (1) and (24) forward in time, with a_1 constant but periodically upgraded to keep it orthonormal to b^1, with time step $\Delta t = \mathcal{O}(1/\mu_1(0))$.

3) When the magnitude of g_f is sufficiently small compared with that of g_s, discard g_f from the system equations and proceed as before with the remnant of g.

There are, of course many variants on this idea. The vector a_1 has been treated as a constant for the purpose of numerical integration because Eq. (26) is easier to use than Eq. (25) for the integration of Eq. (24), but this is not necessary. It may be desirable or even necessary to pick $n > 1$. For example, if several negative eigenvalues all have similar large magnitudes, it may be best to group the corresponding reaction groups as fast. Also, complex conjugate eigenvalues will have to be kept together in pairs. If several reaction groups are kept in g_f, they are ordered according to reaction rate, with a_1, b^1 corresponding to the fastest fast reaction group and a_n, b^n to the slowest fast reaction group.

Examples

Let's first return to example 1, the free linear oscillator with small mass, Eq. (8). Application of CSP to linear time-invariant systems is rather trivial but useful for illustration. For this system, J is constant so that α_i, β^i, and μ_i are constant, which in turn implies that a_i and b^i are constant. Indeed, from Eqs. (24) and (27), a_i and b^i are the right and left eigenvectors of J, respectively:

$$a_i = \alpha_i, \qquad b^i = \beta^i \tag{29}$$

The Jacobian and its eigenvalues for the system of Eq. (8) are

$$J = \begin{pmatrix} -1/\varepsilon & -1/\varepsilon \\ 1 & 0 \end{pmatrix}$$

$$\mu = \frac{-1 \pm \sqrt{1 - 4\varepsilon}}{2\varepsilon}$$

The largest eigenvalue in magnitude is

$$\mu_1 = \frac{-1-\sqrt{1-4\varepsilon}}{2\varepsilon}$$

and, since this is negative, we have a boundary-layer situation. From Eqs. (28) and (29),

$$\boldsymbol{a}_1 = \begin{pmatrix} \mu_1\alpha \\ \alpha \end{pmatrix}, \qquad \boldsymbol{b}^1 = \begin{pmatrix} -\varepsilon\mu_1\beta & \beta \end{pmatrix}$$

where from Eq. (18)

$$\alpha\beta = \frac{1}{1-\varepsilon\mu_1^2}$$

Proceeding with the CSP algorithm, we next use Eq. (19) to compute f^1:

$$f^1 = \mu_1\beta(y_1+y_2) + \beta y_1$$

This allows determination of $\boldsymbol{g}_f$ from Eq. (17),

$$\begin{pmatrix} g_{f_1} \\ g_{f_2} \end{pmatrix} = \frac{1}{1-\varepsilon\mu_1^2}\begin{pmatrix} \mu_1^2 y_1 + \mu_1^2 y_2 + \mu_1 y_1 \\ \mu_1 y_1 + \mu_1 y_2 + y_1 \end{pmatrix}$$

and of $\boldsymbol{g}_s$ from Eq. (20),

$$\begin{pmatrix} g_{s_1} \\ g_{s_2} \end{pmatrix} = \frac{1}{1-\varepsilon\mu_1^2}\begin{pmatrix} -1/\varepsilon y_1 - 1/\varepsilon y_2 - \mu_1 y \\ -\mu_1 y_1 - \mu_1 y_2 - \varepsilon\mu_1^2 y_1 \end{pmatrix}$$

It is easily checked that $\boldsymbol{g} = \boldsymbol{g}_f + \boldsymbol{g}_s$. Expanding $\boldsymbol{g}_s$ shows that

$$\begin{pmatrix} g_{s_1} \\ g_{s_2} \end{pmatrix} = \begin{pmatrix} y_2 + \varepsilon(y_1+3y_2) + \cdots \\ -y_2 + \varepsilon(-y_1+2y_2) + \cdots \end{pmatrix}$$

so that if $\boldsymbol{g}_f$ is discarded, the remaining equations agree with the zero-order outer system, Eq. (11), to zero order. Consequently, a numerical implementation of the CSP algorithm for this problem would capture all of the features of a conventional singular perturbation analysis.

Finally, consider the nonlinear example 2, Eq. (13). The Jacobian and the fast eigenvalue are

$$J = \begin{pmatrix} -2y_1/\varepsilon & -2y_2/\varepsilon \\ 1 & 0 \end{pmatrix}$$

$$\mu_1 = \frac{-y_1 - \sqrt{y_1^2 + 2\varepsilon y_2}}{\varepsilon}$$

both of which are now time-dependent. The eigenvectors of J are

$$\alpha_1 = \begin{pmatrix} \mu_1\alpha \\ \alpha \end{pmatrix}, \qquad \beta^1 = \begin{pmatrix} \varepsilon\mu_1\beta/2y_2 & \beta \end{pmatrix}$$

with

$$\alpha\beta = \frac{1}{1+\varepsilon\mu_1^2/2y_2}$$

According to CSP, we choose at $t=0$,

$$\boldsymbol{a}_1 = \alpha_1(0) = \begin{pmatrix} -\sqrt{(2/\varepsilon)}\,\alpha \\ \alpha \end{pmatrix}$$

$$\boldsymbol{b}^1(0) = \beta^1(0) = \begin{pmatrix} -\sqrt{(\varepsilon/2)}\,\beta & \beta \end{pmatrix}$$

Equations (12) and (24), with Λ_k^i given by Eq. (26), are then integrated forward with $\Delta t = \mathcal{O}(1/\mu_1(0)) = \mathcal{O}(\varepsilon)$, holding $\boldsymbol{a}_1$ constant over each integration step. When $|g_f| \ll |g_s|$, we are in the outer region and $\boldsymbol{g}_f$ is discarded and the integration step size is increased to $\Delta t = \mathcal{O}(1)$.

To verify the performance of the algorithm in the outer region, we note that $\mathrm{d}\boldsymbol{b}^1/\mathrm{d}t$ tends to zero as time is increased.[6] Consequently, we make the approximation $\mathrm{d}\boldsymbol{b}^1/\mathrm{d}t = 0$, giving

$$\Lambda_1^1 = \mu_1, \qquad \boldsymbol{a}_1 = \alpha_1, \qquad \boldsymbol{b}^1 = \beta^1$$

Taking only the leading terms gives the approximations

$$\boldsymbol{a}_1 = \begin{pmatrix} -2y_1\alpha/\varepsilon \\ 0 \end{pmatrix}, \qquad \boldsymbol{b}^1 = \begin{pmatrix} -y_1\beta/y_2 & \beta \end{pmatrix}$$

with

$$\alpha\beta = \frac{\varepsilon y_2}{2y_1^2}$$

Applying the CSP procedure,

$$f^1 = \left(-\frac{y_1\beta}{\varepsilon y_2}\right)(y_2^2 - y_1^2) + \beta y_1$$

$$\begin{pmatrix} g_{f_1} \\ g_{f_2} \end{pmatrix} = \begin{pmatrix} (y_2^2 - y_1^2)/\varepsilon - y_2 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} g_{s_1} \\ g_{s_2} \end{pmatrix} = \begin{pmatrix} y_2 \\ y_1 \end{pmatrix}$$

Comparison with Eq. (15) shows that the method captures the reduced solution in the outer region exactly.

Concluding Remarks

The application of a new method of solution of stiff systems of differential equations to dynamic systems has been reviewed and illustrated with simple examples. It has particular application to complex (nonlinear, high-order) systems for which conventional singular perturbation techniques are inadequate or cumbersome. It does not rely on identification of small parameters or other a priori time-scale information. In addition to providing an efficient numerical integration algorithm, it can reveal the time-scale separation information.

References

[1] Ardema, M. D., *Singular Perturbations in Systems and Control,* International Centre for Mechanical Sciences (CISM) Course and Lectures No. 280. Springer-Verlag, New York, 1983.

[2] Kokotovic, P. V., Khalil, H. K., and O'Reilly, J., *Singular Perturbation Methods in Control: Analysis and Design,* Academic, London, 1986.

[3] Kokotovic, P. V., and Khalil, H. K., *Singular Perturbation in Systems and Control,* IEEE Press, New York, 1986.

[4] Ardema, M. D., and Rajan, N., "Separation of Time Scales in Aircraft Trajectory Optimization," *Journal of Guidance, Control, and Dynamics,* Vol. 8, No. 2, 1985, pp. 275–278.

[5] Lam, S. H., and Goussis, D. A., "Understanding Complex Chemical Kinetics with Computational Singular Perturbations," Rept. 1799-MAE, Department of Mechanical and Aerospace Engineering, Princeton Univ., Princeton, NJ, 1988.

[6] Lam, S. H., and Goussis, D. A., "Basic Theory and Demonstration of Computational Singular Perturbation for Stiff Equations," The 12th IMACS World Congress on Scientific Computation, Paris, France, 1988.

[7] Wasow, W. R., *Asymptotic Expansions for Ordinary Differential Equations,* Interscience, New York, 1965.

[8] Vasileva, A. B., "Asymptotic Behavior of Solutions to Certain Problems Involving Nonlinear Differential Equations Containing a Small Parameter Multiplying the Highest Derivatives," *Russian Math. Surveys,* Vol. 18, 1963, pp. 13–81.