

EFFICIENT COMPUTATION OF SENSITIVITIES FOR ORDINARY DIFFERENTIAL EQUATION BOUNDARY VALUE PROBLEMS*

RADU SERBAN[†] AND LINDA R. PETZOLD[‡]

Abstract. For models described by ordinary differential equation boundary value problems (ODE BVPs), we derive adjoint equations for sensitivity analysis, giving explicit forms for the boundary conditions of the adjoint boundary value problem. The solutions of the adjoint equations are used to efficiently compute gradients of both integral-form and pointwise constraints. Existence and stability results are given for the adjoint system and its numerical solution. The use of the method is demonstrated for a simple example, where it is seen that the method is particularly advantageous for problems with more than a few parameters.

Key words. sensitivity analysis, ODE boundary value problem, adjoint method

AMS subject classifications. 65L10, 65L99

PII. S0036142900376870

1. Introduction. Sensitivity analysis generates essential information for design optimization, parameter estimation, optimal control, data assimilation, process sensitivity, and experimental design. Virtually any scientific or engineering problem can take advantage of sensitivity analysis, for example, problems in chemical engineering applications, multibody mechanical systems, structural engineering, materials science, electric and electronic circuit simulation, and weather prediction models.

There is a large body of work on methods and software for forward sensitivity analysis of initial value problems (IVPs) for ordinary differential equation (ODE) systems [16] and differential-algebraic equation (DAE) systems [11, 22]. Recent research [21, 25] has demonstrated that forward sensitivities can be computed reliably and efficiently via automatic differentiation [4] in combination with ODE/DAE/PDE solution techniques designed to exploit the structure of the sensitivity system.

Forward sensitivity analysis has been shown to be very efficient for problems in which the sensitivities of a (potentially) very large number of quantities, with respect to relatively few parameters, are needed. However, for problems where the number of uncertain parameters is large, the forward sensitivity method becomes computationally intractable. The adjoint (reverse) method is advantageous in the complementary situation, where the sensitivities of a few quantities, with respect to a large number of parameters, are needed. Adjoint sensitivity analysis is particularly attractive for boundary value problems (BVPs). In contrast to the situation for IVPs, where the adjoint method requires considerable memory resources above what is required for the solution of the original problem, the solution values required by

*Received by the editors August 16, 2000; accepted for publication (in revised form) December 19, 2001; published electronically May 1, 2002.

<http://www.siam.org/journals/sinum/40-1/37687.html>

[†]Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94551 (radu@llnl.gov). The research of this author was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory, under contract W-7405-Eng-48.

[‡]Department of Computer Science, University of California, Santa Barbara, CA 93106 (petzold@engineering.ucsb.edu). The research of this author was partially supported by NSF ACI-0086061, DOE DE-FG03-00ER25430, NSF KDI ATM-9873133, NSF-ARPA PC 239415, and EPRI WO-8333-06.

the adjoint method for BVPs are naturally available from the solution of the original problem.

Adjoint sensitivity analysis raises a set of entirely new issues ranging from existence of adjoint operators [3, 5, 6, 20] to construction of adjoint models [10, 17, 18], derivation of boundary conditions for the adjoint states [5], and algorithm implementation [13]. Adjoint sensitivity analysis for BVPs has focused mainly on models described by partial differential equations (PDEs). We cite here the work of Cacuci [5, 6] on general sensitivity theory for nonlinear systems, that of Ghione and Filicori on sensitivity of semiconductor devices [12], the work of Giles and Pierce [14] on adjoint equations in computational fluid dynamics, and that of Machiels, Maclay, and Patera [23, 24] on the use of adjoint methods to obtain a posteriori finite element output bounds.

In [5, 6, 26], adjoint operators are constructed for general nonlinear systems, and results are given for solvability of the original and adjoint systems. However, because of the generality of this setting, boundary conditions for the adjoint problem cannot be explicitly constructed. Instead, for each particular example, proper boundary conditions are obtained by imposing the condition that the Lagrange identity is satisfied. For the adjoint equations for inviscid and viscous compressible flow, Giles and Pierce [14] constructed correct boundary conditions for adjoint problems used in evaluating integral-form quantities. In computational fluid dynamics, most quantities of interest are in integral form. However, in other engineering areas, point quantities such as maximum stresses and/or deformations in structural analysis are of major concern. Being able to efficiently compute gradients of such quantities is thus of high interest.

In the present work we derive adjoint equations for sensitivity analysis of models described by ODE BVPs. For a general form ODE BVP, which is assumed to be well conditioned and to have a unique solution, we derive in section 2 adjoint systems for which we explicitly construct proper boundary conditions. Our goals are to demonstrate that the adjoint method offers an efficient means of computing ODE BVP sensitivities, particularly if there are many parameters, and to show how the adjoint method is formulated for ODE BVPs for different classes of derived functions. Thus we derive adjoint equations to efficiently evaluate not only gradients of integral-form quantities, but also (using the Leibnitz integral rule) gradients of pointwise constraints. In section 3 we establish that, for the problems considered here, the adjoint problems are well-posed and inherit the stability of the original system. We show that numerical stability of the midpoint method for the original system implies numerical stability for the adjoint system. In section 4 we illustrate the computation of sensitivities via the adjoint method on a simple example.

2. Derivation of the adjoint BVP. Consider a state vector $\mathbf{x} \in R^N$ that satisfies the BVP depending on parameters $\mathbf{p} \in R^{N_p}$,

$$(1) \quad \begin{aligned} \mathbf{F}(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{p}, t) &= \mathbf{0}, \\ \mathbf{h}(\mathbf{x}(a), \mathbf{x}(b), \mathbf{p}) &= \mathbf{0}, \end{aligned}$$

and the function $g(\mathbf{x}, \mathbf{p}, t)$ whose gradient with respect to $\mathbf{p}$, $dg/d\mathbf{p}$ is to be evaluated at some time $\tau \in [a, b]$. We assume that the Jacobian of $\mathbf{F}$ with respect to the vector $\dot{\mathbf{x}}$ is nonsingular (meaning that (1) represents a system of ODEs and not DAEs) and that $\mathbf{h}$ represents a set of N independent equations. Note that if g also depends on the time derivatives $\dot{\mathbf{x}}$, then the first set of relations (1) can be used to express g as a function of only $\mathbf{x}$, $\mathbf{p}$, and t . We assume also that (1) is well conditioned and has a unique solution.

In section 2.1 we derive the gradient $(dg/d\mathbf{p})$ at $\tau \in (a, b)$. Using the resulting expressions, the particular case of computing $(dg/d\mathbf{p})$ at $\tau = b$ is analyzed in section 2.2.

2.1. Gradients of g between the integration bounds. We start by deriving the gradient $(dg/d\mathbf{p})$ at some $\tau \in (a, b)$. The derivation closely follows the IVP case [7], with differences arising from the definition of proper boundary conditions for the adjoint equations.

First, define the function

$$G^\tau(\mathbf{p}) = \int_a^\tau g(\mathbf{x}, \mathbf{p}, t) dt.$$

The gradient of G^τ with respect to $\mathbf{p}$ is then simply

$$(2) \quad \frac{dG^\tau}{d\mathbf{p}} = \int_a^\tau \frac{dg}{d\mathbf{p}}(\mathbf{x}, \mathbf{p}, t) dt = \int_a^\tau (g_{\mathbf{p}} + g_{\mathbf{x}}\mathbf{x}_{\mathbf{p}}) dt,$$

where subscripts represent partial differentiation. Applying the Leibnitz integral rule we obtain

$$\frac{d}{d\tau} \frac{dG^\tau}{d\mathbf{p}} = \left. \frac{dg}{d\mathbf{p}} \right|_\tau.$$

Thus $(dg/d\mathbf{p})|_\tau$ can be computed as

$$\left. \frac{dg}{d\mathbf{p}} \right|_\tau = \frac{d}{d\tau} \left(\int_a^\tau (g_{\mathbf{p}} + g_{\mathbf{x}}\mathbf{x}_{\mathbf{p}}) dt \right).$$

The challenge of adjoint sensitivity analysis is now to compute the above quantity without solving for the sensitivities $\mathbf{x}_{\mathbf{p}}$. To do this, we first form the linear sensitivity problem from the BVP (1),

$$(3) \quad \begin{aligned} \mathbf{F}_{\dot{\mathbf{x}}} \dot{\mathbf{x}}_{\mathbf{p}} + \mathbf{F}_{\mathbf{x}} \mathbf{x}_{\mathbf{p}} + \mathbf{F}_{\mathbf{p}} &= \mathbf{0}, \\ \mathbf{A} \mathbf{x}_{\mathbf{p}}(a) + \mathbf{B} \mathbf{x}_{\mathbf{p}}(b) + \mathbf{h}_{\mathbf{p}} &= \mathbf{0}, \end{aligned}$$

where $\mathbf{A} = \mathbf{h}_{\mathbf{x}_0}(\mathbf{x}(a), \mathbf{x}(b), \mathbf{p})$ and $\mathbf{B} = \mathbf{h}_{\mathbf{x}_1}(\mathbf{x}(a), \mathbf{x}(b), \mathbf{p})$. Then, for arbitrary $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in R^N$, the following holds:

$$(4) \quad 0 \equiv \int_a^\tau \boldsymbol{\lambda}_1^* (\mathbf{F}_{\dot{\mathbf{x}}} \dot{\mathbf{x}}_{\mathbf{p}} + \mathbf{F}_{\mathbf{x}} \mathbf{x}_{\mathbf{p}} + \mathbf{F}_{\mathbf{p}}) dt + \int_\tau^b \boldsymbol{\lambda}_2^* (\mathbf{F}_{\dot{\mathbf{x}}} \dot{\mathbf{x}}_{\mathbf{p}} + \mathbf{F}_{\mathbf{x}} \mathbf{x}_{\mathbf{p}} + \mathbf{F}_{\mathbf{p}}) dt,$$

where $*$ indicates the transposed conjugate. Integrating by parts, the first term in the first integral in the above relation becomes

$$\int_a^\tau \boldsymbol{\lambda}_1^* \mathbf{F}_{\dot{\mathbf{x}}} \dot{\mathbf{x}}_{\mathbf{p}} dt = (\boldsymbol{\lambda}_1^* \mathbf{F}_{\dot{\mathbf{x}}} \mathbf{x}_{\mathbf{p}})|_a^\tau - \int_a^\tau \left(\dot{\boldsymbol{\lambda}}_1^* \mathbf{F}_{\dot{\mathbf{x}}} + \boldsymbol{\lambda}_1^* \frac{d\mathbf{F}_{\dot{\mathbf{x}}}}{dt} \right) \mathbf{x}_{\mathbf{p}} dt,$$

where

$$(5) \quad \boldsymbol{\lambda}_1^* \frac{d\mathbf{F}_{\dot{\mathbf{x}}}}{dt} = \left[\left(\frac{d\mathbf{F}_{\dot{\mathbf{x}}}}{dt} \right)^* \boldsymbol{\lambda}_1 \right]^* = (\mathbf{F}_{\dot{\mathbf{x}}}^* \bar{\boldsymbol{\lambda}}_1)_t^* + [(\mathbf{F}_{\dot{\mathbf{x}}}^* \bar{\boldsymbol{\lambda}}_1)_{\mathbf{x}} \dot{\mathbf{x}}]^* + [(\mathbf{F}_{\dot{\mathbf{x}}}^* \bar{\boldsymbol{\lambda}}_1)_{\dot{\mathbf{x}}} \ddot{\mathbf{x}}]^*.$$

A bar over a variable indicates that the variable is held fixed for the purpose of the current differentiation. Without loss of generality, we can assume that $\mathbf{F}$ depends

linearly on $\dot{\mathbf{x}}$ and that therefore, the last term in (5) is zero. Indeed, any other case can be reduced to this one by introducing the additional variables $\mathbf{y} = \dot{\mathbf{x}}$. So from now on, we calculate $\lambda_1^*(d\mathbf{F}_{\dot{\mathbf{x}}}/dt)$ by

$$\lambda_1^* \frac{d\mathbf{F}_{\dot{\mathbf{x}}}}{dt} = (\mathbf{F}_{\dot{\mathbf{x}}}^* \bar{\lambda}_1)_t^* + [(\mathbf{F}_{\dot{\mathbf{x}}}^* \bar{\lambda}_1)_{\mathbf{x}} \dot{\mathbf{x}}]^*.$$

Thus we have from (4) that

$$(6) \quad \begin{aligned} 0 &\equiv (\lambda_1^* \mathbf{F}_{\dot{\mathbf{x}}} \mathbf{x}_{\mathbf{p}})|_a^\tau + (\lambda_2^* \mathbf{F}_{\dot{\mathbf{x}}} \mathbf{x}_{\mathbf{p}})|_\tau^b - \int_a^\tau \left(\dot{\lambda}_1^* \mathbf{F}_{\dot{\mathbf{x}}} - \lambda_1^* \mathbf{F}_{\mathbf{x}} + \lambda_1^* \frac{d\mathbf{F}_{\dot{\mathbf{x}}}}{dt} \right) \mathbf{x}_{\mathbf{p}} dt \\ &\quad - \int_\tau^b \left(\dot{\lambda}_2^* \mathbf{F}_{\dot{\mathbf{x}}} - \lambda_2^* \mathbf{F}_{\mathbf{x}} + \lambda_2^* \frac{d\mathbf{F}_{\dot{\mathbf{x}}}}{dt} \right) \mathbf{x}_{\mathbf{p}} dt + \int_a^\tau \lambda_1^* \mathbf{F}_{\mathbf{p}} dt + \int_\tau^b \lambda_2^* \mathbf{F}_{\mathbf{p}} dt. \end{aligned}$$

A suitable choice for λ_1 and λ_2 to compute $dG^\tau/d\mathbf{p}$ is given by the following.

PROPOSITION 1. *If λ_1 and λ_2 satisfy*

$$(7) \quad \begin{aligned} \dot{\lambda}_1^* \mathbf{F}_{\dot{\mathbf{x}}} - \lambda_1^* \left(\mathbf{F}_{\mathbf{x}} - \frac{d\mathbf{F}_{\dot{\mathbf{x}}}}{dt} \right) &= g_{\mathbf{x}}, \\ \dot{\lambda}_2^* \mathbf{F}_{\dot{\mathbf{x}}} - \lambda_2^* \left(\mathbf{F}_{\mathbf{x}} - \frac{d\mathbf{F}_{\dot{\mathbf{x}}}}{dt} \right) &= \mathbf{0}, \\ \bar{\mathbf{A}} \mathbf{F}_{\dot{\mathbf{x}}}^*(a) \lambda_1(a) + \bar{\mathbf{B}} \mathbf{F}_{\dot{\mathbf{x}}}^*(b) \lambda_2(b) &= \mathbf{0}, \\ \lambda_1(\tau) - \lambda_2(\tau) &= \mathbf{0}, \end{aligned}$$

where $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ are such that

$$(8) \quad \text{span} \begin{bmatrix} \bar{\mathbf{A}}^T \\ \bar{\mathbf{B}}^T \end{bmatrix} = \text{null} [-\mathbf{A}|\mathbf{B}],$$

that is, the rows of $[\bar{\mathbf{A}}|\bar{\mathbf{B}}]$ span the null space of $[-\mathbf{A}|\mathbf{B}]$, then

$$(9) \quad \frac{dG^\tau}{d\mathbf{p}} = -\boldsymbol{\alpha}^* \mathbf{h}_{\mathbf{p}} + \int_a^\tau (g_{\mathbf{p}} + \lambda_1^* \mathbf{F}_{\mathbf{p}}) dt + \int_\tau^b \lambda_2^* \mathbf{F}_{\mathbf{p}} dt,$$

where $\boldsymbol{\alpha} = (\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*)^{-1} (-\mathbf{A}(\mathbf{F}_{\dot{\mathbf{x}}}^* \lambda_1)(a) + \mathbf{B}(\mathbf{F}_{\dot{\mathbf{x}}}^* \lambda_2)(b))$.

Proof. First note that the requirement that the boundary conditions in (1) represent N linearly independent equations is equivalent to $[\mathbf{A}|\mathbf{B}]$ (as well as $[-\mathbf{A}|\mathbf{B}]$) having full row rank. As a consequence, the matrix $\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*$ is invertible. The definition (8) of $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ implies that the rows of $[-\mathbf{A}|\mathbf{B}]$ span the null space of $[\bar{\mathbf{A}}|\bar{\mathbf{B}}]$. On the other hand, the third relation in (7) imposes that the vector $[(\mathbf{F}_{\dot{\mathbf{x}}}^* \lambda_1)(a), (\mathbf{F}_{\dot{\mathbf{x}}}^* \lambda_2)(b)]$ is in the null space of $[\bar{\mathbf{A}}|\bar{\mathbf{B}}]$. Therefore, there exists a vector $\boldsymbol{\alpha} \in R^N$ such that

$$(10) \quad \begin{aligned} (\mathbf{F}_{\dot{\mathbf{x}}}^* \lambda_1)(a) &= -\mathbf{A}^* \boldsymbol{\alpha}, \\ (\mathbf{F}_{\dot{\mathbf{x}}}^* \lambda_2)(b) &= \mathbf{B}^* \boldsymbol{\alpha}, \end{aligned}$$

and thus

$$\begin{aligned} (\lambda_1^* \mathbf{F}_{\dot{\mathbf{x}}} \mathbf{x}_{\mathbf{p}})|_a^\tau + (\lambda_2^* \mathbf{F}_{\dot{\mathbf{x}}} \mathbf{x}_{\mathbf{p}})|_\tau^b &= (\lambda_1(\tau) - \lambda_2(\tau))^* (\mathbf{F}_{\dot{\mathbf{x}}} \mathbf{x}_{\mathbf{p}})(\tau) \\ - (\lambda_1^* \mathbf{F}_{\dot{\mathbf{x}}})(a) \mathbf{x}_{\mathbf{p}}(a) + (\lambda_2^* \mathbf{F}_{\dot{\mathbf{x}}})(b) \mathbf{x}_{\mathbf{p}}(b) &= \boldsymbol{\alpha}^* (\mathbf{A} \mathbf{x}_{\mathbf{p}}(a) + \mathbf{B} \mathbf{x}_{\mathbf{p}}(b)) = -\boldsymbol{\alpha}^* \mathbf{h}_{\mathbf{p}}, \end{aligned}$$

where the second relation in (3) and the last relation in (7) have been used. Since $[\mathbf{A}|\mathbf{B}]$ has full row rank, the $N \times N$ matrix $\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*$ is nonsingular, and $\boldsymbol{\alpha}$ can be computed from (10) as

$$\boldsymbol{\alpha} = (\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*)^{-1} (-\mathbf{A}(\mathbf{F}_{\dot{\mathbf{x}}}^* \boldsymbol{\lambda}_1)(a) + \mathbf{B}(\mathbf{F}_{\dot{\mathbf{x}}}^* \boldsymbol{\lambda}_2)(b)).$$

Substituting this result together with the first two relations of (7) into (6), we have

$$0 = -\boldsymbol{\alpha}^* \mathbf{h}_{\mathbf{p}} - \int_a^\tau g_{\mathbf{x}} \mathbf{x}_{\mathbf{p}} dt + \int_a^\tau \boldsymbol{\lambda}_1^* \mathbf{F}_{\mathbf{p}} dt + \int_\tau^b \boldsymbol{\lambda}_2^* \mathbf{F}_{\mathbf{p}} dt$$

and therefore

$$\frac{dG^\tau}{d\mathbf{p}} = \int_a^\tau (g_{\mathbf{p}} + g_{\mathbf{x}} \mathbf{x}_{\mathbf{p}}) dt = -\boldsymbol{\alpha}^* \mathbf{h}_{\mathbf{p}} + \int_a^\tau (g_{\mathbf{p}} + \boldsymbol{\lambda}_1^* \mathbf{F}_{\mathbf{p}}) dt + \int_\tau^b \boldsymbol{\lambda}_2^* \mathbf{F}_{\mathbf{p}} dt. \quad \square$$

Returning to the problem of computing $dg/d\mathbf{p}$ at τ , by taking the total derivative with respect to τ in (9) we obtain

$$\frac{d}{d\tau} \frac{dG^\tau}{d\mathbf{p}} = \frac{d}{d\tau} \left(-\boldsymbol{\alpha}^* \mathbf{h}_{\mathbf{p}} + \int_a^\tau (g_{\mathbf{p}} + \boldsymbol{\lambda}_1^* \mathbf{F}_{\mathbf{p}}) dt + \int_\tau^b \boldsymbol{\lambda}_2^* \mathbf{F}_{\mathbf{p}} dt \right)$$

and therefore

$$(11) \quad \left. \frac{dg}{d\mathbf{p}} \right|_\tau = -\boldsymbol{\alpha}_\tau^T \mathbf{h}_{\mathbf{p}} + g_{\mathbf{p}}(\tau) + \int_a^\tau \boldsymbol{\mu}_1^* \mathbf{F}_{\mathbf{p}} dt + \int_\tau^b \boldsymbol{\mu}_2^* \mathbf{F}_{\mathbf{p}} dt,$$

where we have used $\boldsymbol{\lambda}_1(\tau) = \boldsymbol{\lambda}_2(\tau)$. The quantities $\boldsymbol{\mu}_1 = (\boldsymbol{\lambda}_1)_\tau$ and $\boldsymbol{\mu}_2 = (\boldsymbol{\lambda}_2)_\tau$ are the solution of the following sensitivity system, obtained by direct differentiation of (7):

$$\begin{aligned} \dot{\boldsymbol{\mu}}_1^* \mathbf{F}_{\dot{\mathbf{x}}} - \boldsymbol{\mu}_1^* \left(\mathbf{F}_{\dot{\mathbf{x}}} - \frac{d\mathbf{F}_{\dot{\mathbf{x}}}}{dt} \right) &= \mathbf{0}, \\ \dot{\boldsymbol{\mu}}_2^* \mathbf{F}_{\dot{\mathbf{x}}} - \boldsymbol{\mu}_2^* \left(\mathbf{F}_{\dot{\mathbf{x}}} - \frac{d\mathbf{F}_{\dot{\mathbf{x}}}}{dt} \right) &= \mathbf{0}, \\ \bar{\mathbf{A}} \mathbf{F}_{\dot{\mathbf{x}}}^*(a) \boldsymbol{\mu}_1(a) + \bar{\mathbf{B}} \mathbf{F}_{\dot{\mathbf{x}}}^*(b) \boldsymbol{\mu}_2(b) &= \mathbf{0}, \\ \boldsymbol{\mu}_1(\tau) + \dot{\boldsymbol{\lambda}}_1(\tau) - \boldsymbol{\mu}_2(\tau) - \dot{\boldsymbol{\lambda}}_2(\tau) &= \mathbf{0}. \end{aligned}$$

The last boundary condition is obtained by taking the total derivative with respect to τ of the boundary condition $\boldsymbol{\lambda}_1(\tau) - \boldsymbol{\lambda}_2(\tau) = \mathbf{0}$ and taking into account all dependencies on τ . These can be better seen if $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are considered as functions of two arguments, $\boldsymbol{\lambda}_1(t, \tau)$ and $\boldsymbol{\lambda}_2(t, \tau)$. In this case, direct differentiation of

$$\boldsymbol{\lambda}(t, \tau)|_{t=\tau} - \boldsymbol{\lambda}(t, \tau)|_{t=\tau} = \mathbf{0}$$

yields

$$\boldsymbol{\lambda}_{1t}(\tau, \tau) + \boldsymbol{\lambda}_{1\tau}(\tau, \tau) - \boldsymbol{\lambda}_{2t}(\tau, \tau) - \boldsymbol{\lambda}_{2\tau}(\tau, \tau) = \mathbf{0},$$

that is,

$$\dot{\boldsymbol{\lambda}}_1(\tau) + \boldsymbol{\mu}_1(\tau) - \dot{\boldsymbol{\lambda}}_2(\tau) + \boldsymbol{\mu}_2(\tau) = \mathbf{0}.$$

Note that, upon substitution of $\dot{\lambda}_1(\tau)$ and $\dot{\lambda}_2(\tau)$, this boundary condition can be further simplified to

$$\mu_1(\tau) - \mu_2(\tau) + (g_x \mathbf{F}_x^{-1})^* (\tau) = \mathbf{0}.$$

The quantity α_τ is obtained by taking the total derivative of α with respect to τ :

$$\alpha_\tau = (\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*)^{-1} (-\mathbf{A}(\mathbf{F}_x^* \mu_1)(a) + \mathbf{B}(\mathbf{F}_x^* \mu_2)(b)).$$

2.2. Gradients of g at the integration bounds. The gradient of $G^b = \int_a^b g(x, p, t) dt$ can be derived by applying a similar procedure, leading to

$$(12) \quad \frac{dG^b}{dp} = -\alpha^* h_p + \int_a^b (g_p + \lambda^* f_p) dt,$$

where λ is the solution of

$$\begin{aligned} \dot{\lambda}^* \mathbf{F}_x - \lambda^* \left(\mathbf{F}_x - \frac{d\mathbf{F}_x}{dt} \right) &= g_x, \\ \bar{\mathbf{A}}\mathbf{F}_x^*(a)\lambda(a) + \bar{\mathbf{B}}\mathbf{F}_x^*(b)\lambda(b) &= \mathbf{0}, \end{aligned}$$

and $\alpha = (\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*)^{-1} (-\mathbf{A}(\mathbf{F}_x^* \lambda)(a) + \mathbf{B}(\mathbf{F}_x^* \lambda)(b)).$

The gradient of g at $t = b$ could, in principle, be obtained as in the previous section by taking the total derivative of (12) with respect to b . Such an approach would be considerably complicated by the fact that g now depends on b implicitly through x . However, if we take

$$\left. \frac{dg}{dp} \right|_b = \lim_{\tau \rightarrow b} \left. \frac{dg}{dp} \right|_\tau,$$

then these difficulties can be circumvented. Indeed, if we specify $\tau = b$ in (11), we obtain

$$(13) \quad \left. \frac{dg}{dp} \right|_b = -\alpha_b^* h_p + g_p(b) + \int_a^b \mu^* \mathbf{F}_p dt,$$

where $\mu = \lambda_b$ is the solution of

$$\begin{aligned} \dot{\mu}^* \mathbf{F}_x - \mu^* \left(\mathbf{F}_x - \frac{d\mathbf{F}_x}{dt} \right) &= \mathbf{0}, \\ \bar{\mathbf{A}}\mathbf{F}_x^*(a)\mu(a) + \bar{\mathbf{B}}\mathbf{F}_x^*(b) \left(\mu(b) + (g_x \mathbf{F}_x^{-1})^* (b) \right) &= \mathbf{0}, \end{aligned}$$

or, rearranging the boundary condition, is the solution of

$$\begin{aligned} \dot{\mu}^* \mathbf{F}_x - \mu^* \left(\mathbf{F}_x - \frac{d\mathbf{F}_x}{dt} \right) &= \mathbf{0}, \\ \bar{\mathbf{A}}\mathbf{F}_x^*(a)\mu(a) + \bar{\mathbf{B}}\mathbf{F}_x^*(b)\mu(b) &= -\bar{\mathbf{B}}g_x^*(b). \end{aligned}$$

The expression of α_b in (13) can be derived as

$$\alpha_b = (\mathbf{A}\mathbf{A}^* + \mathbf{B}\mathbf{B}^*)^{-1} (-\mathbf{A}(\mathbf{F}_x^* \mu)(a) + \mathbf{B}(\mathbf{F}_x^* \mu + g_x^*)(b)).$$

3. On existence and stability of the adjoint solution. Consider a linear implicit ODE BVP of the form (3), written here as

$$(14) \quad \begin{aligned} \mathbf{M}(t)\dot{\mathbf{x}} + \mathbf{K}(t)\mathbf{x} + \mathbf{q}(t) &= \mathbf{0}, \\ \mathbf{A}\mathbf{x}(a) + \mathbf{B}\mathbf{x}(b) + \mathbf{c} &= \mathbf{0}, \end{aligned}$$

whose adjoint BVP can be written as

$$(15) \quad \begin{aligned} \frac{d}{dt}(\mathbf{M}^*(t)\boldsymbol{\lambda}) - \mathbf{K}^*(t)\boldsymbol{\lambda} + \mathbf{r}(t) &= \mathbf{0}, \\ \bar{\mathbf{A}}\mathbf{M}^*(a)\boldsymbol{\lambda}(a) + \bar{\mathbf{B}}\mathbf{M}^*(b)\boldsymbol{\lambda}(b) &= \mathbf{0}, \end{aligned}$$

where $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ are defined as before.

In this section we investigate the stability of the adjoint system. More precisely, if the original system is stable, will the adjoint system also be stable? If we consider the adjoint system (15), the answer may be negative. Indeed, consider the following IVP example [7]:

$$(16) \quad e^t \dot{x} + \frac{1}{2}e^t x = 0,$$

with some initial condition at $t = a$. This system is equivalent to

$$\dot{x} + \frac{1}{2}x = 0,$$

so it is stable to integration from the left. However, the adjoint system (15) for (16) is

$$(17) \quad e^t \dot{\lambda} - \frac{1}{2}e^t \lambda + e^t \lambda = 0 \quad \Rightarrow \quad \dot{\lambda} + \frac{1}{2}\lambda = 0.$$

Note that the adjoint system must be solved in a backward direction. Thus the adjoint system (17) is unstable.

Denoting $\bar{\boldsymbol{\lambda}}(t) = \mathbf{M}^*(t)\boldsymbol{\lambda}(t)$, we can form the *augmented adjoint system* for (15),

$$(18) \quad \begin{aligned} \dot{\bar{\boldsymbol{\lambda}}} - \mathbf{K}^*(t)\boldsymbol{\lambda} + \mathbf{r}(t) &= \mathbf{0}, \\ \bar{\boldsymbol{\lambda}} - \mathbf{M}^*(t)\boldsymbol{\lambda} &= \mathbf{0}, \\ \bar{\mathbf{A}}\mathbf{M}^*(a)\boldsymbol{\lambda}(a) + \bar{\mathbf{B}}\mathbf{M}^*(b)\boldsymbol{\lambda}(b) &= \mathbf{0}. \end{aligned}$$

If, instead of (17), we solve the augmented adjoint system (18), then $\bar{\boldsymbol{\lambda}}$ satisfies

$$\dot{\bar{\boldsymbol{\lambda}}} - \frac{1}{2}\bar{\boldsymbol{\lambda}} = 0,$$

which is stable to integration from the right. We will show that, in general, if the original system (14) is stable, then the augmented adjoint system (18) for $\bar{\boldsymbol{\lambda}}$ is stable. First, note that since $\mathbf{M}$ is nonsingular, $\bar{\boldsymbol{\lambda}}$ satisfies

$$\begin{aligned} \dot{\bar{\boldsymbol{\lambda}}} - \mathbf{K}^*(t)(\mathbf{M}^*(t))^{-1}\bar{\boldsymbol{\lambda}} + \mathbf{r}(t) &= \mathbf{0}, \\ \bar{\mathbf{A}}\bar{\boldsymbol{\lambda}}(a) + \bar{\mathbf{B}}\bar{\boldsymbol{\lambda}}(b) &= \mathbf{0}. \end{aligned}$$

In other words, for the implicit ODE BVP, the augmented adjoint system for $\bar{\boldsymbol{\lambda}}$ is the same as the adjoint system of the explicit ODE BVP equivalent to the original

system (14). Therefore it is sufficient to investigate stability of the adjoint system for the explicit ODE BVP,

$$(19) \quad \begin{aligned} \dot{\mathbf{x}} &= \mathbf{C}(t)\mathbf{x} + \mathbf{q}(t), \\ \mathbf{A}\mathbf{x}(a) + \mathbf{B}\mathbf{x}(b) &= \mathbf{c}, \end{aligned}$$

whose adjoint BVP can be written as

$$(20) \quad \begin{aligned} \dot{\boldsymbol{\lambda}} &= -\mathbf{C}^*(t)\boldsymbol{\lambda} + \mathbf{r}(t), \\ \bar{\mathbf{A}}\boldsymbol{\lambda}(a) + \bar{\mathbf{B}}\boldsymbol{\lambda}(b) &= \mathbf{0}. \end{aligned}$$

We begin by deriving the relationship between fundamental solutions of these two problems. This is given by the following.

LEMMA 1. *Let $\mathbf{X}$ and $\boldsymbol{\Lambda}$ be any fundamental solutions of (19) and (20), respectively. Then, for any $t, s \in [a, b]$, $\boldsymbol{\Lambda}^*(t)\mathbf{X}(t) = \boldsymbol{\Lambda}^*(s)\mathbf{X}(s)$.*

Proof. Consider $\mathbf{Z}(t) = \boldsymbol{\Lambda}^*(t)\mathbf{X}(t)$. Then

$$\dot{\mathbf{Z}} = \dot{\boldsymbol{\Lambda}}^*(t)\mathbf{X}(t) + \boldsymbol{\Lambda}^*(t)\dot{\mathbf{X}}(t) = (-\mathbf{C}^*(t)\boldsymbol{\Lambda}(t))^*\mathbf{X}(t) + \boldsymbol{\Lambda}^*(t)(\mathbf{C}(t)\mathbf{X}(t)) = \mathbf{0}.$$

Therefore $\boldsymbol{\Lambda}^*(t)\mathbf{X}(t) = \boldsymbol{\Lambda}^*(s)\mathbf{X}(s)$ for all $t, s \in [a, b]$; in particular, $\boldsymbol{\Lambda}^*(t)\mathbf{X}(t) = \boldsymbol{\Lambda}^*(a)\mathbf{X}(a) = \boldsymbol{\Lambda}^*(b)\mathbf{X}(b)$ for any $t \in [a, b]$. $\square$

As a direct consequence of Lemma 1, $\|\boldsymbol{\Lambda}(t)\boldsymbol{\Lambda}^{-1}(s)\| = \|\mathbf{X}(s)\mathbf{X}^{-1}(t)\|$ for all $s \geq t$. This proves the following.

THEOREM 1. *The adjoint system (20) of an (asymptotically) stable linear ODE IVP (19) is (asymptotically) stable.*

We now concentrate on the ODE BVP. We first show the following.

THEOREM 2. *If the BVP (19) has a unique solution, then a solution for the adjoint BVP (20) exists and is unique.*

Proof. Consider the fundamental solution $\mathbf{X}(t)$ of the homogeneous equivalent of (19) which satisfies $\mathbf{X}(a) = \mathbf{I}$. Then the matrix $\mathbf{Q} = \mathbf{A} + \mathbf{B}\mathbf{X}(b)$ is nonsingular [2]. Similarly, let $\boldsymbol{\Lambda}(t)$ be the fundamental solution of the homogeneous equivalent of (20), which satisfies $\boldsymbol{\Lambda}(b) = \mathbf{I}$, and construct the matrix $\bar{\mathbf{Q}} = \bar{\mathbf{A}}\boldsymbol{\Lambda}(a) + \bar{\mathbf{B}}$. From Lemma 1 we have that $\boldsymbol{\Lambda}(a) = \mathbf{X}^*(b)$. Then

$$\mathbf{Q}\bar{\mathbf{A}}^* = \mathbf{A}\bar{\mathbf{A}}^* + \mathbf{B}\mathbf{X}(b)\bar{\mathbf{A}}^* = \mathbf{B}\bar{\mathbf{B}}^* + \mathbf{B}\boldsymbol{\Lambda}^*(a)\bar{\mathbf{A}}^* = \mathbf{B}\bar{\mathbf{Q}}^*$$

and

$$\begin{aligned} \mathbf{Q}\mathbf{X}^{-1}(b)\bar{\mathbf{B}}^* &= \mathbf{A}\mathbf{X}^{-1}(b)\bar{\mathbf{B}}^* + \mathbf{B}\bar{\mathbf{B}}^* = \mathbf{A}(\boldsymbol{\Lambda}^*(a))^{-1}\bar{\mathbf{B}}^* + \mathbf{A}\bar{\mathbf{A}}^* \\ &= \mathbf{A}(\boldsymbol{\Lambda}^*(a))^{-1}\bar{\mathbf{Q}}^* = \mathbf{A}\mathbf{X}^{-1}(b)\bar{\mathbf{Q}}^*, \end{aligned}$$

where we have used $\mathbf{A}\bar{\mathbf{A}}^* = \mathbf{B}\bar{\mathbf{B}}^*$. Since $\mathbf{Q}$ is invertible, we can write

$$\begin{aligned} \bar{\mathbf{A}}^* &= \mathbf{Q}^{-1}\mathbf{B}\bar{\mathbf{Q}}^*, \\ \bar{\mathbf{B}}^* &= \mathbf{X}(b)\mathbf{Q}^{-1}\mathbf{A}\mathbf{X}^{-1}(b)\bar{\mathbf{Q}}^*. \end{aligned}$$

Thus

$$\begin{bmatrix} \bar{\mathbf{A}}^* \\ \bar{\mathbf{B}}^* \end{bmatrix} = \begin{bmatrix} \mathbf{Q}^{-1}\mathbf{B} \\ \mathbf{X}(b)\mathbf{Q}^{-1}\mathbf{A}\mathbf{X}^{-1}(b) \end{bmatrix} \bar{\mathbf{Q}}^*.$$

Since $[\bar{\mathbf{A}}|\bar{\mathbf{B}}]$ has full row rank, it follows that $\bar{\mathbf{Q}}^*$ has full rank. Therefore, $\bar{\mathbf{Q}}$ is invertible and (20) has a unique solution. $\square$

Stability results for the adjoint problem are given by the following [9].

THEOREM 3. *If the BVP (19) is well conditioned, then its adjoint BVP (20) is well conditioned.*

We now consider numerical stability for the adjoint system. As a consequence of Theorem 3, zero-stability (i.e., stability as the stepsize $h \rightarrow 0$ and the number of steps $n \rightarrow \infty$) for the adjoint BVP (20) is inherited from zero-stability of the original BVP (19). Here we are concerned with the question of whether a numerical method which is stable for the original system (19) on the mesh

$$\pi : a = t_0 < t_1 < \cdots < t_{N-1} < t_N = b$$

is also stable for the adjoint system (20). We consider the midpoint method for which we show the following.

THEOREM 4. *Numerical stability of the midpoint method for the original system on some mesh π implies numerical stability for the adjoint system on the same mesh.*

Proof. Discretizing the original system (19) with the midpoint rule, we obtain

$$(21) \quad \begin{aligned} \frac{\mathbf{x}_n - \mathbf{x}_{n-1}}{h_n} &= \mathbf{C}(t_{n-1/2}) \frac{\mathbf{x}_n + \mathbf{x}_{n-1}}{2} + \mathbf{q}(t_{n-1/2}), \quad n = 1, \dots, N, \\ \mathbf{A}\mathbf{x}_0 + \mathbf{B}\mathbf{x}_N &= -\mathbf{c}, \end{aligned}$$

where $h_n = t_n - t_{n-1}$. The first N relations in (21) can be written as

$$-\mathbf{S}_n \mathbf{x}_{n-1} + \mathbf{R}_n \mathbf{x}_n = \mathbf{q}(t_{n-1/2}), \quad n = 1, \dots, N,$$

where

$$\begin{aligned} \mathbf{S}_n &= \frac{1}{h_n} \mathbf{I} + \frac{1}{2} \mathbf{C}(t_{n-1/2}), \\ \mathbf{R}_n &= \frac{1}{h_n} \mathbf{I} - \frac{1}{2} \mathbf{C}(t_{n-1/2}). \end{aligned}$$

Thus we have that $[\mathbf{x}_0^*; \mathbf{x}_N^*]$ is the solution of a linear system of the form

$$(22) \quad \begin{bmatrix} \mathbf{P} & -\mathbf{I} \\ \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{q}} \\ -\mathbf{c} \end{bmatrix}$$

for some right-hand side $\hat{\mathbf{q}}$, where $\mathbf{P} = \mathbf{R}_N^{-1} \mathbf{S}_N \mathbf{R}_{N-1}^{-1} \mathbf{S}_{N-1} \cdots \mathbf{R}_1^{-1} \mathbf{S}_1$.

The midpoint method applied to the adjoint problem (20) yields the linear equations

$$\mathbf{S}_n^* \boldsymbol{\lambda}_n - \mathbf{R}_n^* \boldsymbol{\lambda}_{n-1} = \mathbf{r}(t_{n-1/2}), \quad n = 1, \dots, N.$$

Since $\mathbf{S}_n$ and $\mathbf{R}_n^{-1}$ commute, it follows that $[\boldsymbol{\lambda}_N^*; \boldsymbol{\lambda}_0^*]$ is the solution of a linear system

$$(23) \quad \begin{bmatrix} \mathbf{P}^* & -\mathbf{I} \\ \mathbf{B} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_N \\ \boldsymbol{\lambda}_0 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{r}} \\ \mathbf{0} \end{bmatrix}$$

for some right-hand side $\hat{\mathbf{r}}$. We show next that with $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ defined by (8), solving linear system (23) is equivalent to solving linear system (22). Indeed, system (22) can be solved as

$$\begin{aligned} (\mathbf{A} + \mathbf{B}\mathbf{P})\mathbf{x}_0 &= \mathbf{B}\hat{\mathbf{q}} - \mathbf{c}, \\ \mathbf{x}_N &= \mathbf{P}\mathbf{x}_0 - \hat{\mathbf{q}}. \end{aligned}$$

On the other hand, by construction, there exists a vector $\boldsymbol{\alpha}$ defined as in (10). Thus system (23) can be solved as

$$\begin{aligned} (\mathbf{P}^*\mathbf{B}^* + \mathbf{A}^*)\boldsymbol{\alpha} &= \hat{\mathbf{r}}, \\ \boldsymbol{\lambda}_0 &= -\mathbf{A}^*\boldsymbol{\alpha}, \\ \boldsymbol{\lambda}_N &= \mathbf{B}^*\boldsymbol{\alpha}. \end{aligned}$$

Noting that $\|(\mathbf{P}^*\mathbf{B}^* + \mathbf{A}^*)^{-1}\| = \|(\mathbf{A} + \mathbf{B}\mathbf{P})^{-1}\|$, this concludes the proof. $\square$

4. Numerical example. As an example we consider the following ODE BVP:

$$\begin{aligned} (J + ml^2)\ddot{\theta} &= u(t) - mgl \cos(\theta), \\ \theta(a) &= \theta_0, \\ \theta(b) &= \theta_1, \end{aligned} \tag{24}$$

which describes the motion of a 2-D pendulum of length $2l$, mass m , and inertia J under the action of gravity (g) and a time varying applied torque $u(t)$. The position of the pendulum is imposed at the initial and final times. Considering the torque $u(t)$ parameterized by $p \in R^{N_p}$, we wish to estimate the sensitivity with respect to p of the energy $g(\theta, \dot{\theta}, p, t) = \frac{1}{2}(J + ml^2)\dot{\theta}^2 + mgl \sin(\theta)$ at some time $t \in (a, b)$, as well as the sensitivity of the total energy $G^\tau(p) = \int_a^\tau g(\theta, \dot{\theta}, p, t)dt$ over the interval $[a, \tau]$.

As an alternative to using adjoint sensitivity analysis for the solution of these problems, one could generate the sensitivity ODE BVP systems (3) by the following forward method: For each of the parameters p_i , compute the sensitivities of the trajectories $(\theta(t), \dot{\theta}(t))$ and then, using the chain rule of differentiation, evaluate the gradients of g and G^τ . However, such an approach is computationally expensive, especially if the dimension N_p of the parameterization of $u(t)$ is very large.

We first transform (24) into a first order ODE BVP:

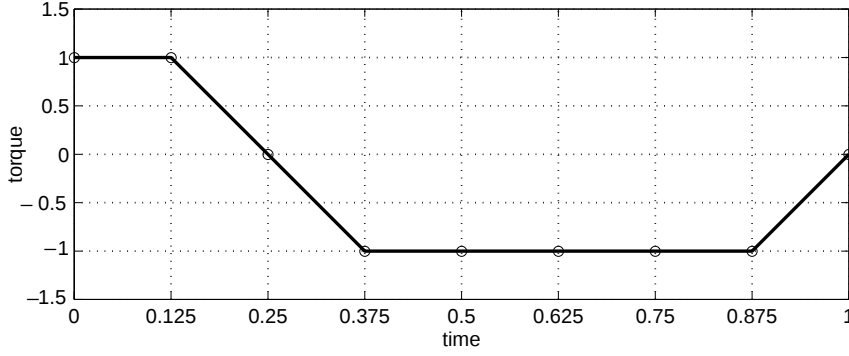
$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= \frac{1}{J + ml^2} (u(t) - mgl \cos(x_1)), \\ x_1(a) &= \theta_0, \\ x_1(b) &= \theta_1, \end{aligned} \tag{25}$$

in which case

$$g(x, p, t) = \frac{1}{2}(J + ml^2)x_2^2 + mgl \sin(x_1) \tag{26}$$

and

$$G^\tau(p) = \int_a^\tau \left(\frac{1}{2}(J + ml^2)x_2^2 + mgl \sin(x_1) \right) dt. \tag{27}$$

FIG. 1. Torque $u(t)$ for the pendulum example ($N = 8$, $N_p = 9$).

Consider a piecewise linear approximation of u given by

$$(28) \quad u(t) = -p_k \frac{t - k\Delta t}{\Delta t} + p_{k+1} \frac{t - (k-1)\Delta t}{\Delta t};$$

$$(k-1)\Delta t \leq t \leq k\Delta t, \quad k = 1, 2, \dots, N,$$

where $\Delta t = (b-a)/N$. This gives $N_p = N + 1$ parameters p . Let $N = 8$ and let u be as in Figure 1. For $a = 0$, $b = 1$, $\theta_0 = 0$, $\theta_1 = 0$, $\tau = 0.25$, and $m = g = l = J = 1$, we compare gradients of g and G^τ at both $\tau \in (a, b)$ and $\tau = b$ obtained by the adjoint sensitivity analysis presented in the previous sections with gradients computed through forward sensitivity analysis. Differences in gradients computed with the two methods are summarized in Table 1.

TABLE 1
Differences in gradients computed with adjoint (a) and forward methods (f).

i	$\left[\frac{dG^\tau}{dp}\right]_{(a)} - \left[\frac{dG^\tau}{dp}\right]_{(f)}$	$\left[\frac{dg(\tau)}{dp}\right]_{(a)} - \left[\frac{dg(\tau)}{dp}\right]_{(f)}$
1	$4.068486 \cdot 10^{-06}$	$-5.219818 \cdot 10^{-09}$
2	$8.042969 \cdot 10^{-06}$	$-9.258004 \cdot 10^{-12}$
3	$7.934972 \cdot 10^{-06}$	$-3.032900 \cdot 10^{-11}$
4	$5.331125 \cdot 10^{-06}$	$1.268874 \cdot 10^{-06}$
5	$9.902579 \cdot 10^{-08}$	$3.172192 \cdot 10^{-06}$
6	$-4.355139 \cdot 10^{-08}$	$5.555800 \cdot 10^{-11}$
7	$-2.891162 \cdot 10^{-08}$	$2.859599 \cdot 10^{-11}$
8	$-1.443129 \cdot 10^{-08}$	$8.619001 \cdot 10^{-12}$
9	$-3.178720 \cdot 10^{-10}$	$5.191848 \cdot 10^{-09}$
i	$\left[\frac{dG^b}{dp}\right]_{(a)} - \left[\frac{dG^b}{dp}\right]_{(f)}$	$\left[\frac{dg(b)}{dp}\right]_{(a)} - \left[\frac{dg(b)}{dp}\right]_{(f)}$
1	$1.013121 \cdot 10^{-05}$	$-4.789156 \cdot 10^{-09}$
2	$1.990550 \cdot 10^{-05}$	$-8.493997 \cdot 10^{-12}$
3	$1.943275 \cdot 10^{-05}$	$-2.782700 \cdot 10^{-11}$
4	$1.906737 \cdot 10^{-05}$	$-5.279502 \cdot 10^{-11}$
5	$1.883816 \cdot 10^{-05}$	$-7.485801 \cdot 10^{-11}$
6	$1.875241 \cdot 10^{-05}$	$-8.563900 \cdot 10^{-11}$
7	$1.879674 \cdot 10^{-05}$	$-7.838902 \cdot 10^{-11}$
8	$1.893979 \cdot 10^{-05}$	$-4.928702 \cdot 10^{-11}$
9	$9.530949 \cdot 10^{-06}$	$-1.530841 \cdot 10^{-06}$

All ODE BVPs involved in both adjoint and forward computations were numerically solved with COLSYS [1]. We note that the version of COLSYS that we used has a limit of 20 on the number of differential equations, thus limiting the number of parameters that we could include for the forward sensitivity system to $N_p = 9$. Of course, when using the adjoint approach this is not an issue, as only one additional BVP of the same size as the original problem must be solved to evaluate gradients with respect to an array of parameters of any size. The other obvious advantage of using adjoint sensitivity versus forward sensitivity is, of course, computational efficiency. Indeed, solution of the BVP (original + adjoint) required by the adjoint approach was 15 times faster than solution of the BVP (original + N_p forward sensitivities) required by the forward approach. In all fairness, we must note that a careful implementation of forward sensitivity analysis (which takes advantage of the special structure of the sensitivity systems and the fact that they share the same Jacobian matrices with the original BVP) will lead to a speedup of only about $(1 + N_p)/(1 + 1) = 5$. Since COLSYS does not provide a sensitivity analysis capability, the overhead computations (especially in the linear algebra) explain the much higher speedup obtained.

REFERENCES

- [1] U.M. ASCHER, J. CHRISTIANSEN, AND R.D. RUSSEL, *A collocation solver for mixed order systems of boundary value problems*, Math. Comp., 33 (1979), pp. 659–679.
- [2] U.M. ASCHER, R.M.M. MATTHEIJ, AND R.D. RUSSELL, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*, SIAM, Philadelphia, 1995.
- [3] K. BALLA AND R. MARZ, *Linear differential algebraic equations of index 1 and their adjoint equations*, Results Math., 37 (2000), pp. 13–35.
- [4] C. BISCHOF, A. CARLE, G. CORLISS, A. GRIEWANK, AND P. HOVLAND, *ADIFOR - Generating derivative codes from Fortran programs*, Scientific Programming, 1 (1992), pp. 11–29.
- [5] D.G. CACUCI, *Sensitivity theory for nonlinear systems. I. Nonlinear functional analysis approach*, J. Math. Phys., 22 (1981), pp. 2794–2802.
- [6] D.G. CACUCI, *Sensitivity theory for nonlinear systems. II. Extension to additional classes of responses*, J. Math. Phys., 22 (1981), pp. 2803–2812.
- [7] Y. CAO, S. LI, L.R. PETZOLD, AND R. SERBAN, *Adjoint sensitivity analysis for differential-algebraic equations: The adjoint DAE system and its numerical solution*, submitted.
- [8] M. CARACOTSIOS AND W.E. STEWART, *Sensitivity analysis of initial value problems with mixed ODEs and algebraic equations*, Computers and Chemical Engineering, 9 (1985), pp. 359–365.
- [9] E.A. CODDINGTON AND N. LEVINSON, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.
- [10] R.M. ERRICO, *What is an adjoint model?* Bull. Amer. Meteorological Society, 78 (1997), pp. 2577–2591.
- [11] W.F. FEEHERY AND P.I. BARTON, *Efficient sensitivity analysis of large-scale differential-algebraic systems*, Appl. Numer. Math., 25 (1997), pp. 41–54.
- [12] G. GHIONE AND F. FILICORI, *A computationally efficient unified approach to the analysis of the sensitivity and noise of semiconductor devices*, IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems, 12 (1993), pp. 425–438.
- [13] R. GIERING AND T. KAMINSKI, *Recipes for adjoint code construction*, ACM Trans. Math. Software, 24 (1998), pp. 437–474.
- [14] M.B. GILES AND N.A. PIERCE, *Adjoint Equations in CFD: Duality, Boundary Conditions and Solution Behaviour*, AIAA Paper 97-1850, American Institute of Aeronautics and Astronautics, New York, 1997.
- [15] G.H. GOLUB AND C.F. VAN LOAN, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1996.
- [16] E. HAIRER, S.P. NORSETT, AND G. WANNER, *Solving Ordinary Differential Equations I*, Springer-Verlag, Berlin, 1987.
- [17] E.J. HAUG AND P.E. EHLE, *Second-order design sensitivity analysis of mechanical system dynamics*, Internat. J. Numer. Methods Engrg., 18 (1982), pp. 1699–1717.

- [18] E.J. HAUG, R. WEHAGE, AND N.C. BARMAN, *Design sensitivity analysis of planar mechanism and machine dynamics*, Trans. ASME, 103 (1981).
- [19] H.B. KELLER, *Numerical Methods for Two-Point Boundary-Value Problems*, Blaisdel, Waltham, MA, 1968.
- [20] R.M. LEWIS, *Numerical Computation of Sensitivities and the Adjoint Approach*, Tech. report 97-61, ICASE, NASA Langley Research Center, Hampton, VA, 1997.
- [21] S. LI AND L.R. PETZOLD, *Design of New DASPK for Sensitivity Analysis*, Tech. report, Department of Computer Science, University of California, Santa Barbara, CA, 1999.
- [22] S. LI AND L.R. PETZOLD, *Software and algorithms for sensitivity analysis of large-scale differential-algebraic systems*, J. Comput. Appl. Math., 125 (2000), pp. 131–145.
- [23] L. MACHIELS, *A Posteriori Finite Element Output Bounds for Discontinuous Galerkin Discretizations of Parabolic Problems*, Tech. report UCRL-JC-136614, Lawrence Livermore National Laboratory, Livermore, CA, 1999.
- [24] L. MACHIELS, Y. MADAY, AND A.T. PATERA, *Output Bounds for Reduced-Order Approximations of Elliptic Partial Differential Equations*, Tech. report FML 99-5-1, Massachusetts Institute of Technology, Cambridge, MA, 1999.
- [25] T. MALY AND L.R. PETZOLD, *Numerical methods and software for sensitivity analysis of differential-algebraic systems*, Appl. Numer. Math., 20 (1997), pp. 57–79.
- [26] G.I. MARCHUK, V.I. AGOSHKOV, AND V.P. SHUTYAEV, *Adjoint Equations and Perturbation Algorithms*, CRC Press, Boca Raton, FL, 1996.