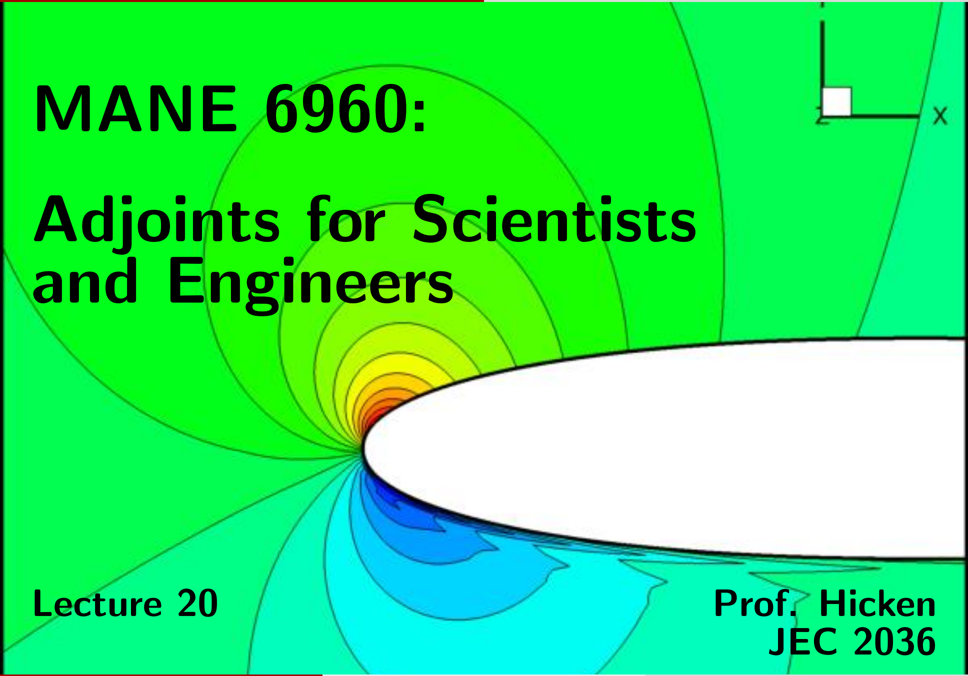

MANE 6960:

Adjoints for Scientists and Engineers

Lecture 20

Prof. Hicken
JEC 2036

Overview

As with adjoint BVPs, one can discretize adjoint IBVPs directly; this is the unsteady version of the continuous-adjoint method.

However, for the same reasons discussed earlier in the context of BVPs, it is often advantageous to consider the discrete-adjoint approach for time-marching methods.

- In the discrete adjoint approach to unsteady problems, we first discretize the primal IBVP and functional, in both space and in time.
- Then we derive a discrete adjoint equation corresponding to these discretized quantities.

Overview (cont.)

We are interested in answering the following questions in the context of the discrete adjoint for time-marching methods:

1. What is the impact on the discrete adjoint if we discretize the state equation using an explicit/implicit time-marching scheme?
2. Is a particular time-marching method adjoint consistent?

Discrete Adjoint of Time Marching Methods

Model Problem

We will make a few simplifying assumptions in order to focus our effort on the time discretization:

1. We will assume the IBVP is linear and autonomous/time-invariant; and
2. We will assume a method-of-lines approach, in which the spatial discretization has already been performed.

We will discuss nonlinear IBVPs later.

It is worth noting that the method-of-lines approach is not the only way to discretize IBVP: we could also use a space-time discretization in which both the spatial and temporal operators are discretized simultaneously.

Model Problem (cont.)

Based on the above assumptions, we can consider the following model initial value problem (IVP):

$$\frac{du_h}{dt} = A_h u_h, \qquad \forall t \in [0, T], \tag{$\star$}$$

$$u_h(0) = u_h^{(0)}, \tag{IC}$$

- Here, $u_h(t) \in \mathbb{R}^s$ is an s-vector corresponding to the spatial degrees of freedom
- $A_h \in \mathbb{R}^{s \times s}$ corresponds to the spatial discretization.

Model Problem (cont.)

For later, it is worth recalling that the solution to the above linear time-invariant system is

$$u_h(t) = e^{A_h t} u_h^{(0)}.$$

- $e^{A_h t}$ is the matrix exponential, defined by

$$e^{A_h t} = \sum_{k=0}^{\infty} \frac{1}{k!} A_h^k t^k$$

Model Problem (cont.)

We will also consider the following functional for the case studies below:

$$J_h(u_h(t)) = \int_0^T g_h(t)^T u_h(t)\, dt + g_T^T u_h(T).$$

- We assume that $g_h(t)$ and g_T incorporate the necessary data for the spatial inner products $(\cdot,\cdot)_\Omega$ and $(\cdot,\cdot)_\Gamma$.

Adjoint of the Model Problem

Before proceeding, we need to derive the adjoint IVP.

- As usual in the linear case, we can do this by subtracting the adjoint-weighted residual and then rearranging to get J to be independent of $u_h(t)$.

$$J_h(u_h(t)) = \int_0^T g_h(t)^T u_h(t)\,dt + g_T^T u_h(T) - \int_0^T \psi_h^T \left(\frac{du_h}{dt} - A_h u_h\right) dt$$

$$= \int_0^T g_h^T u_h\,dt + g_T^T u_h(T) + \int_0^T u_h^T \frac{\partial \psi_h}{\partial t}\,dt + \int_0^T u_h^T A_h^T \psi_h\,dt - \left(u_h^T \psi_h\right)_{t=0}^{t=T}$$

$$= (u_h^{(0)})^T \psi_h(0) - u_h^T(T)\left[\psi_h(T) - g_T\right] - \int_0^T u_h^T\left(-\frac{\partial \psi_h}{\partial t} - A_h^T \psi_h - g_h\right) dt$$

Adjoint of the Model Problem (cont.)

Thus, the adjoint IVP is

$$-\frac{d\psi_h}{dt} = A_h^T \psi_h + g_h, \qquad \forall t \in [0, T], \tag{ADJ}$$

$$\psi_h(T) = g_T, \tag{TC}$$

- As with the primal IVP, the solution to the adjoint IVP can be written in terms of the matrix exponential:

$$\psi_h = e^{A_h^T (T-t)} g_T - \int_T^t e^{-A_h^T (t-\tau)} g(\tau)\, d\tau.$$

Case Study #1: Explicit Midpoint Method

Our first case study uses the explicit midpoint method to discretize the state IVP. This method can be written as a predictor-corrector scheme with two stages as follows:

$$\begin{aligned} \hat{u}_h^{(n+1/2)} &= u_h^{(n)} + \frac{\Delta t}{2} A_h u_h^{(n)}, \qquad && \forall n = 0, 1, 2, \ldots, N-1. \\ u_h^{(n+1)} &= u_h^{(n)} + \Delta t A_h \hat{u}_h^{(n+1/2)}, \qquad && \forall n = 0, 1, 2, \ldots, N-1. \end{aligned}$$

- $\Delta t \equiv T/N$, where N is the number of steps taken.
- The above scheme is a second-order explicit Runge-Kutta scheme.

Case Study #1: Explicit Midpoint Method (cont.)

It is instructive to express $u_h^{(n+1)}$ explicitly in terms of $u_h^{(n)}$, which we can easily do for this linear IVP:

$$u_h^{(n+1)} = u_h^{(n)} + \Delta t A_h \hat{u}_h^{(n+1/2)}$$

$$= u_h^{(n)} + \Delta t A_h \left(u_h^{(n)} + \frac{\Delta t}{2} A_h u_h^{(n)} \right)$$

$$= \underbrace{\left(I + \Delta t A_h + \frac{\Delta t^2}{2} A_h^2 \right)}_{\text{trucated matrix exponential, } e^{\Delta t A_h}} u_h^{(n)}$$

Case Study #1: Explicit Midpoint Method (cont.)

We will use the midpoint quadrature rule to discretize the functional:

$$J_{h,\Delta t} = \sum_{n=0}^{N-1} \frac{\Delta t}{2} \left(g^{(n+1/2)}\right)^T \left(u_h^{(n)} + u_h^{(n+1)}\right) + g_T^T u_h^{(N)}$$

- $g^{(n+1/2)} \equiv g\left((n+1/2)\Delta t\right)$
- For constant $g(t)$, this is equivalent to trapezoid quadrature.
- Other choices are possible, but there is little point in using a quadrature more accurate than second-order, given that this is the temporal accuracy of $u_h^{(n)}$.

Case Study #1: Explicit Midpoint Method (cont.)

Next, we introduce a discrete Lagrangian, and differentiate with respect to $u_h^{(n)}$ and $\hat{u}_h^{(n+1/2)}$ to find the adjoint equations.

$$
\begin{aligned}
L_{h,\Delta t} = \sum_{n=0}^{N-1} \frac{\Delta t}{2} \left(g^{(n+1/2)}\right)^T \left(u_h^{(n)} + u_h^{(n+1)}\right) + g_T^T u_h^{(N)} \\
- \sum_{n=0}^{N-1} \left(\hat{\psi}_h^{(n+1/2)}\right)^T \left[\hat{u}_h^{(n+1/2)} - u_h^{(n)} - \frac{\Delta t}{2} A_h u_h^{(n)}\right] \\
- \sum_{n=0}^{N-1} \left(\psi_h^{(n+1)}\right)^T \left[u_h^{(n+1)} - u_h^{(n)} - \Delta t A_h \hat{u}_h^{(n+1/2)}\right]
\end{aligned}
$$

$$\frac{\partial L_{h,\Delta t}}{\partial \hat{u}_h^{(n+1/2)}} = 0 \quad , \quad \frac{\partial L_{h,\Delta t}}{\partial u_h^{(n)}} = 0$$

Case Study #1: Explicit Midpoint Method (cont.)

Differentiating with respect to $\hat{u}_h^{(n+1/2)}$ we get

$$-\hat{\psi}_h^{(n+1/2)} + \Delta t \, A_h^T \psi_h^{(n+1)} = 0$$

$$\Rightarrow \quad \hat{\psi}_h^{(n+1/2)} = \Delta t \, A_h^T \psi_h^{(n+1)}$$

Case Study #1: Explicit Midpoint Method (cont.)

Differentiating with respect to $u_h^{(n)}$, $n \neq N$, we get

$$\frac{\Delta t}{2}\left(g^{(n+1/2)} + g^{(n-1/2)}\right) + \hat{\psi}_h^{(n+1/2)} + \frac{\Delta t}{2} A_h^T \hat{\psi}^{(n+1/2)} + \psi_h^{(n+1)} - \psi_h^{(n)} = 0$$

$$\Rightarrow \quad \psi_h^{(n)} = \psi_h^{(n+1)} + \hat{\psi}_h^{(n+1/2)} + \frac{\Delta t}{2} A_h^T \hat{\psi}^{(n+1/2)} + \frac{\Delta t}{2}\left(g^{(n+1/2)} + g^{(n-1/2)}\right)$$

Case Study #1: Explicit Midpoint Method (cont.)

The last time step must be considered separately:

$$\frac{\Delta t}{2} g^{(N-1/2)} + g_T - \psi_h^{(N)} = 0$$

$$\Rightarrow \quad \psi_h^{(N)} = g_T + \frac{\Delta t}{2} g^{(N-1/2)}$$

Case Study #1: Explicit Midpoint Method (cont.)

Consider the first question we are interested in answering:

- What is the impact on the discrete adjoint of discretizing the state equation using an explicit time-marching scheme?

Since we get an explicit formula for $\psi_h^{(n)}$ in terms of "earlier" time steps/stages, we see that the discrete adjoint is explicit in this case.

- This is true more generally: explicit-time marching schemes produce explicit discrete adjoint schemes.

Why?

$$\left[\, D_t \,\right] u_h^{(\cdot)} = \left[\, L \,\right] u_h^{(\cdot)}$$

$$\left[\, D_t^T \,\right] \psi_h^{(\cdot)} = \left[\, L^T \,\right] \psi_h^{(\cdot)}$$

Case Study #1: Explicit Midpoint Method (cont.)

Next, we consider the second question:

- Is this particular time-marching method adjoint consistent?

First, consider the equation for $\psi_h^{(N)}$:

$$\psi_h^{(N)} = g_T + \frac{\Delta t}{2} g^{(N-1/2)}$$

$$\lim_{\Delta t \to 0} \psi_h^{(N)} = g_T$$

✓ consistent with (TC)

Case Study #1: Explicit Midpoint Method (cont.)

Next, consider the equation for $\psi_h^{(n)}$, where $n \neq N$:

$$\psi_h^{(n)} = \psi_h^{(n+1)} - \hat{\psi}_h^{(n+1/2)} + \frac{\Delta t}{2} A_h^T \hat{\psi}_h^{(n+1/2)} + \frac{\Delta t}{2}\left(g^{(n+1/2)} + g^{(n-1/2)}\right)$$

$$= \underbrace{\psi_h^{(n+1)} - \Delta t A_h^T \psi_h^{(n+1)} + \frac{\Delta t^2}{2}(A_h^T)^2 \psi_h^{(n+1)}}_{\text{truncated matrix exponential}} + \frac{\Delta t}{2}\left(\quad \right)$$

$$\tilde{\psi}_h^{(n+1/2)} \equiv \psi^{(n+1)} - \frac{\Delta t}{2} A_h^T \psi_h^{(n+1)}$$

$$\psi_h^{(n)} = \psi_h^{(n+1)} - \Delta t A_h^T \tilde{\psi}_h^{(n+1/2)} + \frac{\Delta t}{2}\left(g^{(n+1/2)} + g^{(n-1/2)}\right)$$

Case Study #1: Explicit Midpoint Method (cont.)

The explicit-midpoint method (RK2) is an adjoint consistent method when the functional is discretized using the midpoint quadrature rule.

$O(\Delta t)$

Case Study #2: BDF2

Our second case study looks at the second-order backward difference formula (BDF2). This scheme uses two previous time steps, so the first step needs a different method: we use implicit Euler here.

$$\begin{aligned} u_h^{(1)} &= u_h^{(0)} + \Delta t A_h u_h^{(n)}, \\ 3u_h^{(n+1)} &= 4u_h^{(n)} - u_h^{(n-1)} + 2\Delta t A_h u_h^{(n+1)}, \qquad \forall n = 1, 2, \ldots, N-1. \end{aligned}$$

- As before, $\Delta t \equiv T/N$, where N is the number of steps taken.
- Note that the BDF2 scheme is implicit, since we must invert A_h to find $u_h^{(n+1)}$.

Case Study #2: BDF2 (cont.)

For the functional, we adopt the trapezoid rule:

$$J_{h,\Delta t} = \sum_{n=0}^{N} w_n \Delta t \left(g^{(n)}\right)^T u_h^{(n)} + g_T^T u_h^{(N)}$$

where the trapezoid weights are

$$w_n = \begin{cases} 1, & \forall\, n = 1, 2, \ldots, N-1 \\ \frac{1}{2}, & n = 0, N \end{cases}$$

- $g^{(n)} \equiv g\left(n\Delta t\right)$
- As before, other choices of discretization are possible.

Case Study #2: BDF2 (cont.)

The Lagrangian corresponding to the BDF2 primal discretization and trapezoid functional is

$$
\begin{aligned}
L_{h,\Delta t} = \sum_{n=0}^{N} w_n \Delta t \left(g^{(n)}\right)^T u_h^{(n)} + g_T^T u_h^{(N)} \\
- \sum_{n=1}^{N-1} \left(\psi_h^{(n+1)}\right)^T \left[3u_h^{(n+1)} - 4u_h^{(n)} + u_h^{(n-1)} - 2\Delta t A_h u_h^{(n+1)}\right] \\
- \left(\psi_h^{(1)}\right)^T \left[u_h^{(1)} - u_h^{(0)} - \Delta t A_h u_h^{(n)}\right]
\end{aligned}
$$

In order to derive the discrete adjoint equations, it is helpful to express the Lagrangian in matrix form, as shown on the next slide:

Case Study #2: BDF2 (cont.)

$$L_{h,\Delta t} = \begin{bmatrix} w_1 \Delta t g^{(1)} \\ w_2 \Delta t g^{(2)} \\ \vdots \\ w_N \Delta t g^{(N)} + g_T \end{bmatrix}^T \begin{bmatrix} u_h^{(1)} \\ u_h^{(2)} \\ \vdots \\ u_h^{(N)} \end{bmatrix}$$

$$- \begin{bmatrix} \psi_h^{(1)} \\ \psi_h^{(2)} \\ \psi_h^{(3)} \\ \vdots \\ \psi_h^{(N)} \end{bmatrix}^T \left(\begin{bmatrix} 1 & & & & & \\ -4 & 3 & & & & \\ 1 & -4 & 3 & & & \\ & 1 & -4 & 3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -4 & 3 \end{bmatrix} - 2\Delta t \begin{bmatrix} \frac{1}{2} A_h & & & & \\ & A_h & & & \\ & & A_h & & \\ & & & \ddots & \\ & & & & A_h \end{bmatrix} \right) \begin{bmatrix} u_h^{(1)} \\ u_h^{(2)} \\ u_h^{(3)} \\ \vdots \\ u_h^{(N)} \end{bmatrix}$$

$+$ (terms independent of $u_h^{(n)}$, $n = 1, 2, \ldots, N$)

Case Study #2: BDF2 (cont.)

Differentiating with respect to $u_h^{(n)}$, $n = 2, 3, \ldots, N-2$, we get

$$w_n \Delta t\, g^{(n)} - 3\psi_h^{(n)} + 4\psi_h^{(n+1)} + 2\Delta t\, A_h^T \psi_h^{(n)} - \psi_h^{(n+2)} = 0$$

$$\Rightarrow \quad 3\psi_h^{(n)} = 4\psi_h^{(n+1)} - \psi_h^{(n+2)} + 2\Delta t\, A_h^T \psi_h^{(n)} + \Delta t\, w_n g^{(n)}$$

Aside:

$$\frac{1}{2\Delta t}\left(3\psi^{(n)} - 4\psi^{(n+1)} + \psi_h^{(n+2)}\right) = -\frac{\partial \psi}{\partial t} + O(\Delta t^2)$$

Case Study #2: BDF2 (cont.)

Differentiating with respect to the last state, $u_h^{(N)}$, we get

$$w_N \Delta t \, g^{(N)} + g_T - 3\psi_h^{(N)} + 2\Delta t A_h^T \psi_h^{(N)} = 0$$

$$\Rightarrow \quad (3I - 2\Delta t A_h^T)\psi_h^{(N)} = g_T + w_N \Delta t \, g^{(N)}$$

Case Study #2: BDF2 (cont.)

Differentiating with respect to the second-last state, $u_h^{(N-1)}$, we get

$$w_{N-1}\,\Delta t\, g^{(N-1)} - 3\psi_h^{(N-1)} + 4\psi_h^{(N)} + 2\Delta t\, A_h^T \psi_h^{(N-1)} = 0$$

$$\Rightarrow \quad 3\psi_h^{(N-1)} = 4\psi_h^{(N)} + 2\Delta t\, A_h^T \psi_h^{(N-1)} + w_{N-1}\,\Delta t\, g^{(N-1)}$$

Case Study #2: BDF2 (cont.)

Finally, differentiating with respect to the first state, $u_h^{(1)}$, we get

$$w_1 \Delta t\, g^{(1)} - \psi_h^{(1)} + 4\psi_h^{(2)} - \psi_h^{(3)} + \Delta t\, A_h^T \psi_h^{(1)} = 0$$

$$\Rightarrow \quad \psi_h^{(1)} = 4\psi_h^{(2)} - \psi_h^{(3)} + \Delta t\, A_h^T \psi_h^{(1)} + w_1 \Delta t\, g^{(1)}$$

Case Study #2: BDF2 (cont.)

As before, consider the first question:

- What is the impact on the discrete adjoint of discretizing the state equation using an implicit time-marching scheme?

Since we get a linear system for $\psi_h^{(n)}$, we see that the discrete adjoint is implicit in this case.

- This is true more generally: implicit-time marching schemes produce implicit discrete adjoint schemes.

Case Study #2: BDF2 (cont.)

Now, the second question:

- Is this particular time-marching method adjoint consistent?

We begin with the interior adjoints, $\psi_h^{(n)}$, $n = 2, 3, \ldots, N-2$:

$$\underbrace{\frac{3\psi_h^{(n)} - 4\psi_h^{(n+1)} + \psi_h^{(n+2)}}{2\Delta t}}_{-\left.\frac{\partial \psi}{\partial t}\right|_{t=t^{(n)}} + O(\Delta t^2)} = A_h^T \psi_h^{(n)} + \frac{1}{2} g^{(n)}$$

factor of 1/2 makes this adjoint inconsistent

(Can correct by defining $\tilde{\psi}_h^{(n)} = 2\psi_h^{(n)}$)

Case Study #2: BDF2 (cont.)

For the last adjoint, we have

$$(3I - 2\Delta t\, A_h^T)\,\psi_h^{(N)} = g_T + \overset{\frac{1}{2}}{\downarrow}{w_N}\, \Delta t\, g^{(N)}$$

$$\lim_{\Delta t \to 0} 3\psi_h^{(N)} = g_T \qquad \left(\text{want } \psi(T) = g_T\right)$$

Could introduce a scaled adjoint

$$\tilde{\psi}_h^{(N)} = 3\psi_h^{(N)}$$

but this would be inconsistent with previously proposed scaling.

Case Study #2: BDF2 (cont.)

For the second-last adjoint, we find

$$\underbrace{3\psi_h^{(N-1)} = 4\psi_h^{(N)}}_{} + 2\Delta t A_h^T \psi_h^{(N-1)} + w_{N-1}\Delta t g^{(N-1)}$$

Cannot turn this into $\frac{\partial \psi}{\partial t} + O(\Delta t^p)$

Case Study #2: BDF2 (cont.)

And, for the first adjoint we have

$$\underbrace{\psi_h^{(1)} = 4\psi_h^{(2)} - \psi_h^{(3)}} + \Delta t A_h^T \psi_h^{(1)} + w_1 \Delta t g^{(1)}$$

$$\frac{1}{\Delta t}\left(\psi_h^{(1)} - 4\psi_h^{(2)} + \psi_h^{(3)}\right) = \frac{1}{\Delta t}\Big(\psi_h^{(1)} - 4\psi_h^{(1)} - 4\frac{\partial \psi}{\partial t}\Delta t + \psi_h^{(1)} + 2\frac{\partial \psi}{\partial t}\Delta t + O(\Delta t^2)\Big)$$

$$= \frac{-2}{\Delta t}\psi_h^{(1)} - 2\left.\frac{\partial \psi}{\partial t}\right|_{t=t^{(1)}} + O(\Delta t)$$

or, try to extrapolate

$$\psi_h^{(1)} = 4\psi_h^{(2)} - \psi^{(3)} + O(\Delta t)$$

$$\psi_h^{(1)} = 2\psi_h^{(1)} + O(\Delta t)$$!! (But scaling might help)

Case Study #2: BDF2 (cont.)

Therefore,

The BDF2 method is an adjoint inconsistent method.

References

References