Sensitivity of differential-algebraic equations

Yingbo Ma
mayingbo5@gmail.com

March 12, 2020

1 Notation

- $M \in \mathbb{C}^{m\times m}$ denotes a mass matrix.
- $t_0 < t_1 \in \mathbb{C}$ denote constants.
- $t \in \mathbb{C}$ denotes independent variable.
- $p \in \mathbb{C}^n$ denotes parameters.
- $u(p,t) : \mathbb{C}^n \times \mathbb{C} \mapsto \mathbb{C}^m$ denotes states.
- $f(u(p,t),p,t) : \mathbb{C}^m \times \mathbb{C}^n \times \mathbb{C} \mapsto \mathbb{C}^m$ denotes the right-hand-side of a differential equation.
- $S(p,t) : \mathbb{C}^n \times \mathbb{C} \mapsto \mathbb{C}^{m\times n}$ denotes $\frac{\partial u}{\partial p}$ (sensitivity with respect to the states).
- $J(u(p,t),p,t) : \mathbb{C}^m \times \mathbb{C}^n \times \mathbb{C} \mapsto \mathbb{C}^{m\times m}$ denotes $\frac{\partial f}{\partial u}$ (Jacobian of the differential equation with respect to the states).
- $g(u(p,t),p,t) : \mathbb{C}^m \times \mathbb{C}^n \times \mathbb{C} \mapsto \mathbb{C}$ is a cost function that is sufficiently smooth.
- $\lambda(u(p,t),p,t,t_0,t_1) : \mathbb{C}^m \times \mathbb{C}^n \times \mathbb{C} \times \mathbb{C} \times \mathbb{C} \mapsto \mathbb{C}^n$ denotes Lagrange multiplier.
- λ_τ denotes $\frac{\partial \lambda}{\partial \tau}$.

2 Introduction and Forward Sensitivity Analysis

In many applications, we may wish to compute the gradient of the continuous cost function

$$G[u] = \int_{t_0}^{t_1} g(u(p,t),p,t)\,\mathrm{d}t \tag{1}$$

with respect to the parameters $\frac{\mathrm{d}G}{\mathrm{d}p}$, where u is a function that satisfies the differential-algebraic equation

$$M\dot{u} = f(u(p,t), p, t), \tag{2}$$
$$u(t_0) = u_0(p). \tag{3}$$

Naïvely, we could apply Leibniz rule for integration and obtain:

$$\begin{align}
\frac{\mathrm{d}G}{\mathrm{d}p} &= \frac{\mathrm{d}}{\mathrm{d}p}\int_{t_0}^{t_1} g(u(p,t),p)\,\mathrm{d}t \tag{4}\\
&= \int_{t_0}^{t_1} \frac{\mathrm{d}}{\mathrm{d}p} g(u(p,t),p)\,\mathrm{d}t \tag{5}\\
&= \int_{t_0}^{t_1} \frac{\partial g}{\partial u}\frac{\partial u}{\partial p} + \frac{\partial g}{\partial p}\,\mathrm{d}t \tag{6}\\
&= \int_{t_0}^{t_1} \frac{\partial g}{\partial u} S + \frac{\partial g}{\partial p}\,\mathrm{d}t \tag{7}
\end{align}$$

We can get S by differentiating eq. (2) with respect to p both sides,

$$\begin{align}
\frac{\mathrm{d}}{\mathrm{d}p}\frac{\mathrm{d}}{\mathrm{d}t} Mu = \frac{\mathrm{d}}{\mathrm{d}p} f(u(p,t),p,t) \tag{8}\\
\implies M\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}u}{\mathrm{d}p} = \frac{\partial f}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}p} + \frac{\partial f}{\partial p} \tag{9}\\
\implies M\dot{S} = JS + \frac{\partial f}{\partial p}. \tag{10}
\end{align}$$

eq. (10) is often referred as the forward sensitivity equation. It is apparent that computing S can be intractable when there are large number of parameters, because we would have to solve a $m \times n$ differential equation.

3 Adjoint Sensitivity Analysis with Continuous Cost Function

To alleviate the computational cost of the forward sensitivity equation, we could add a "0" to $\frac{\mathrm{d}G}{\mathrm{d}p}$, and get:

$$\begin{align}
\frac{\mathrm{d}G}{\mathrm{d}p} =& \frac{\mathrm{d}I}{\mathrm{d}p} \tag{11}\\
=& \int_{t_0}^{t_1} \frac{\partial g}{\partial u} S + \frac{\partial g}{\partial p} - \lambda^* \left(\underbrace{M\dot{S} - JS - \frac{\partial f}{\partial p}}_{0 \quad eq.\ (10)} \right) \mathrm{d}t\,. \tag{12}
\end{align}$$

The motivation of this step is to introduce $\dot{S}$ and an ***arbitrary*** λ function, and the hope is to use the classic technique of integration by parts to group the S terms and choose the λ function such that the gradient is ***independent*** of S.

After integration by parts, the $\lambda^* M\dot{S}$ term is

$$\int_{t_0}^{t_1} \lambda^* M\dot{S}\,\mathrm{d}t = \lambda^* MS\Big|_{t_0}^{t_1} - \int_{t_0}^{t_1} \dot{\lambda}^* MS\,\mathrm{d}t\,. \tag{13}$$

The gradient expression after grouping is then:

$$\begin{aligned}
\frac{\mathrm{d}G}{\mathrm{d}p} &= \int_{t_0}^{t_1} \frac{\partial g}{\partial u}S + \frac{\partial g}{\partial p} + \dot{\lambda}^* MS + \lambda^* JS + \lambda^* \frac{\partial f}{\partial p}\,\mathrm{d}t - \lambda^* MS\Big|_{t_0}^{t_1} \\
&= \int_{t_0}^{t_1} \left(\frac{\partial g}{\partial u} + \dot{\lambda}^* M + \lambda^* J\right)S + \lambda^* \frac{\partial f}{\partial p} + \frac{\partial g}{\partial p}\,\mathrm{d}t - \lambda^* MS\Big|_{t_0}^{t_1}
\end{aligned} \tag{14}$$

It becomes obvious that we can impose the condition

$$\frac{\partial g}{\partial u} + \dot{\lambda}^* M + \lambda^* J = \mathbf{0} \quad \text{and} \quad \lambda(t_1)^* M = \mathbf{0}^*, \tag{15}$$

to make eq. (14) independent of S.

After rearranging eq. (14) and eq. (15), we have

$$M^*\dot{\lambda} = -J^*\lambda - \frac{\partial g}{\partial u}^*, \tag{16}$$

$$M^*\lambda(t_1) = \mathbf{0} \tag{17}$$

$$\frac{\mathrm{d}G}{\mathrm{d}p} = \int_{t_0}^{t_1} \lambda^* \frac{\partial f}{\partial p} + \frac{\partial g}{\partial p}\,\mathrm{d}t + \lambda^* MS\Big|_{t=t_0}. \tag{18}$$

Here, we want to remark that the $\lambda^* MS\big|_{t=t_0}$ term is zero if the initial condition of eq. (2) is independent of the parameters.

4 Adjoint Sensitivity Analysis with Discrete Cost Function

In many cases, we only want to compute sensitivity at specific a time point τ, i.e. $\frac{\partial g(u(p,t),p,t)}{\partial p}\Big|_{t=\tau}$. In these cases, we call the cost function discrete.

To reuse our previous result, we need to find a way to relate $\frac{\mathrm{d}g}{\mathrm{d}p}$ to $\frac{\mathrm{d}G}{\mathrm{d}p}$. There are two options, using Dirac delta distribution or using Leibniz integration rule. We pick the latter because it is more straight forward.

Note we have

$$\frac{\mathrm{d}}{\mathrm{d}p}\frac{\mathrm{d}}{\mathrm{d}\tau}G[u] = \frac{\mathrm{d}}{\mathrm{d}p}\frac{\mathrm{d}}{\mathrm{d}\tau}\int_{t_0}^{\tau} g(u(p,t),p,t)\,\mathrm{d}t \tag{19}$$

$$= \frac{\mathrm{d}}{\mathrm{d}p}\left[g(u,p,\tau) + \int_{t_0}^{\tau} \frac{\partial g}{\partial \tau}\,\mathrm{d}t\right] \tag{20}$$

$$= \frac{\mathrm{d}g}{\mathrm{d}p} + \int_{t_0}^{\tau} \frac{\partial^2 g}{\partial \tau \partial p}\,\mathrm{d}t\,, \tag{21}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\mathrm{d}}{\mathrm{d}p}G[u] = \frac{\mathrm{d}}{\mathrm{d}\tau}\left[\int_{t_0}^{\tau} \lambda^* \frac{\partial f}{\partial p} + \frac{\partial g}{\partial p}\,\mathrm{d}t + \lambda^* MS\Big|_{t=t_0}\right] \quad \text{eq. (18)} \tag{22}$$

$$= \left(\lambda^* \frac{\partial f}{\partial p} + \frac{\partial g}{\partial p}\right)\Big|_{t=\tau} + \int_{t_0}^{\tau} \lambda_\tau^* \frac{\partial f}{\partial p} + \underbrace{\lambda^* \frac{\partial^2 f}{\partial p \partial \tau}}_{=0}\,\mathrm{d}t + \int_{t_0}^{\tau} \frac{\partial^2 g}{\partial \tau \partial p}\,\mathrm{d}t + \lambda_\tau^* MS\Big|_{t=t_0}. \tag{23}$$

By comparing eq. (21) and eq. (23), we can conclude that

$$\frac{\mathrm{d}g}{\mathrm{d}p} = \left(\lambda^* \frac{\partial f}{\partial p} + \frac{\partial g}{\partial p}\right)\Big|_{t=\tau} + \int_{t_0}^{\tau} \lambda_\tau^* \frac{\partial f}{\partial p}\,\mathrm{d}t + \lambda_\tau^* MS\Big|_{t=t_0}. \tag{24}$$

Now we need λ_τ to compute the sensitivity integral eq. (24). It can be obtained by differentiating eq. (16),

$$M^* \dot{\lambda_\tau} = -J^* \lambda_\tau - \frac{\partial g}{\partial \tau} \tag{25}$$

$$= -J^* \lambda_\tau \quad g \text{ does not depend on } \tau \quad . \tag{26}$$

We want to remark that from eq. (26), we can see that the initialization of λ_τ cannot be trivial, since $\lambda_\tau(\tau) = \mathbf{0}$ can only result in uninteresting dynamics.

It is important to note that λ dependents on not only t but also τ with a discrete cost function, since λ_τ is non-zero.

To obtain the initialization of $\lambda_\tau\big|_{t=\tau}$, we can differentiate $\lambda(t,\tau)\big|_{t=\tau}$ in the constraint eq. (17) with respect to τ,

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\lambda(t,\tau)^* M\Big|_{t=\tau}\right) = \left(\underbrace{\frac{\partial \lambda}{\partial t}\frac{\partial t}{\partial \tau}\Big|_{t=\tau}}_{\dot{\lambda}}\right)^* M + \lambda_\tau^* M\Big|_{t=\tau} = \mathbf{0}^*. \tag{27}$$

Hence, we have

$$\lambda_\tau^* M\Big|_{t=\tau} = -\,\dot{\lambda}^* M\Big|_{t=\tau} \tag{28}$$

$$= \left(J^* \lambda + \frac{\partial g}{\partial u}^*\right)^*\Big|_{t=\tau} \quad \text{eq. (16)} \quad . \tag{29}$$

Together, the system of equations is,

$$M^* \dot{\lambda_\tau} = -J^* \lambda_\tau \tag{30}$$

$$M^* \lambda_\tau\Big|_{t=\tau} = \left(J^* \lambda + \frac{\partial g}{\partial u}^*\right)\Big|_{t=\tau} \tag{31}$$

$$\frac{\mathrm{d}g}{\mathrm{d}p} = \left(\lambda^* \frac{\partial f}{\partial p} + \frac{\partial g}{\partial p}\right)\Big|_{t=\tau} + \int_{t_0}^{\tau} \lambda_\tau^* \frac{\partial f}{\partial p}\,\mathrm{d}t + \lambda_\tau^* MS\Big|_{t=t_0}. \tag{32}$$

4.1 Example: Index-I differential-algebraic equation

To handle singular mass matrix, we need to split the problem system into differential variables $\{\cdot\}^d$ and algebraic variables $\{\cdot\}^a$, so that $\widetilde{M}$ is fully ranked,

$$\begin{pmatrix} \widetilde{M} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \dot{u}^d \\ \dot{u}^a \end{pmatrix} = \begin{pmatrix} f(u^d, u^a, p, t) \\ h(u^d, u^a, p, t) \end{pmatrix} \tag{33}$$

$$u^d(t_0) = u_{d0}(p). \tag{34}$$

We have

$$J^* = \begin{pmatrix} \frac{\partial f}{\partial u^d}^* & \frac{\partial h}{\partial u^d}^* \\ \frac{\partial f}{\partial u^a}^* & \frac{\partial h}{\partial u^a}^* \end{pmatrix}. \tag{35}$$

The initialization step is

$$\begin{pmatrix} \widetilde{M}^* & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda_\tau^d \\ \lambda_\tau^a \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial u^d}^* & \frac{\partial h}{\partial u^d}^* \\ \frac{\partial f}{\partial u^a}^* & \frac{\partial h}{\partial u^a}^* \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \lambda^a \end{pmatrix} + \begin{pmatrix} \frac{\partial g^d}{\partial u}^* \\ \frac{\partial g^a}{\partial u}^* \end{pmatrix}. \tag{36}$$

Expanding the equation we have

$$\widetilde{M}^* \lambda_\tau^d = \frac{\partial h}{\partial u^d}^* \lambda^a + \frac{\partial g^d}{\partial u}^* \tag{37}$$

$$\mathbf{0} = \frac{\partial h}{\partial u^a}^* \lambda^a + \frac{\partial g^a}{\partial u}^* \implies \lambda^a = -\left(\frac{\partial h}{\partial u^a}^*\right)^{-1} \frac{\partial g^a}{\partial u}^* \tag{38}$$

Plugging eq. (38) into eq. (37), we obtain the initialization of λ_τ

$$\widetilde{M}^* \lambda_\tau^d(\tau) = \left(-\frac{\partial h}{\partial u^d}^* \left(\frac{\partial h}{\partial u^a}^*\right)^{-1} \frac{\partial g^a}{\partial u}^* + \frac{\partial g^d}{\partial u}^* \right)\Bigg|_{t=\tau}. \tag{39}$$