Adjoint formulation for manifold-reduced system

Seung Whan Chung

January 9, 2017

Original governing equation

We refer to the original governing equation as a priori projected system:

$$\begin{aligned}
\frac{d\mathbf{x}}{dt} &= \mathbf{P}(\mathbf{x};\theta)\cdot\mathbf{F}(\mathbf{x};\theta)\\
\mathbf{P}(\mathbf{x};\theta) &= \mathbf{I} - \mathbf{V}_f\mathbf{U}_f\\
\mathbf{J} = \nabla_x\mathbf{F} &= \begin{bmatrix}\mathbf{V}_s & \mathbf{V}_f\end{bmatrix}\begin{bmatrix}\Lambda_{ss} & 0\\ 0 & \Lambda_{ff}\end{bmatrix}\begin{bmatrix}\mathbf{U}_s\\ \mathbf{U}_f\end{bmatrix}
\end{aligned}\tag{1}$$

where $\theta \in \mathbb{R}^{N_c}$ is the control parameter. For other notations, refer to the original document.

Observable and its sensitivity

We have the observable as a functional of state variable $\mathbf{x}$:

$$\mathcal{O}[\mathbf{x}] = \int_0^T o(\mathbf{x})dt,$$

and we would like to measure its sensitivity to control parameters,

$$\begin{aligned}
\nabla_\theta\mathcal{O} &= \int_0^T \nabla_x o\cdot\nabla_\theta\mathbf{x}\; dt + \int_0^T \nabla_\theta o\; dt\\
&= \langle\nabla_x o, \nabla_\theta\mathbf{x}\rangle + \int_0^T \nabla_\theta o\; dt,
\end{aligned}\tag{2}$$

where the inner product is defined on $\mathbb{R}^N$,

$$\langle\mathbf{a},\mathbf{b}\rangle = \int_0^T a_i b_i\; dt$$

with Einstein notation for indices.

Sensitivity equation

To obtain the sensitivity (2), the sensitivity of the solution $\nabla_\theta\mathbf{x}$ needs to be computed. It can be obtained by solving the linearized equation from (1). From now on the equations are expressed in index notation to avoid the ambiguity.

- i, j, k: the index on $\mathbb{R}^N$ space
- c: the index on control space $\mathbb{R}^{N_c}$
- l: the index on fast manifold space $\mathbb{R}^{N_f}$

$$
\begin{aligned}
\frac{d}{dt}\partial_c x_i &= \partial_c \left(P_{ij}F_j\right) \\
&= \partial_k \left(P_{ij}F_j\right)\partial_c x_k + \partial_c \left(P_{ij}F_j\right) \\
&= \left[\partial_k P_{ij}\cdot F_j + P_{ij}\cdot \partial_k F_j\right]\partial_c x_k + \partial_c P_{ij}F_j + P_{ij}\partial_c F_j
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
\partial_k P_{ij} &= -\partial_k V_{il}U_{lj} - V_{il}\partial_k U_{lj} \\
\partial_c P_{ij} &= -\partial_c V_{il}U_{lj} - V_{il}\partial_c U_{lj}
\end{aligned}
\tag{4}
$$

Adjoint formulation

We define a constraint functional $\mathcal{H}$ as the inner product of adjoint variable and (3),

$$
\begin{aligned}
\mathcal{H} &= \left\langle \mathbf{x}^\dagger, \frac{d}{dt}\partial_c x_i - \left[\partial_k P_{ij}\cdot F_j + P_{ij}\cdot \partial_k F_j\right]\partial_c x_k - \partial_c P_{ij}F_j - P_{ij}\partial_c F_j \right\rangle \\
&\equiv 0.
\end{aligned}
\tag{5}
$$

Since $\mathcal{H}$ is always equivalent to 0 as (3) holds, we add (5) to the sensitivity (2),

$$
\nabla_\theta \mathcal{O} = \langle \nabla_x o, \nabla_\theta \mathbf{x}\rangle + \int_0^T \nabla_\theta o \; dt = \langle \nabla_x o, \nabla_\theta \mathbf{x}\rangle + \mathcal{H} + \int_0^T \nabla_\theta o \; dt
$$

which can be expressed in index-wise notation,

$$
\begin{aligned}
\nabla_\theta \mathcal{O} &= \langle \nabla_x o, \nabla_\theta \mathbf{x}\rangle + \mathcal{H} + \int_0^T \nabla_\theta o \; dt \\
&= \langle \partial_i o, \partial_c x_i\rangle + \left\langle x_i^\dagger, \frac{d}{dt}\partial_c x_i\right\rangle - \left\langle x_i^\dagger, \left[\partial_k P_{ij}\cdot F_j + P_{ij}\cdot \partial_k F_j\right]\partial_c x_k\right\rangle \\
&\qquad - \left\langle x_i^\dagger, \partial_c P_{ij}F_j\right\rangle - \left\langle x_i^\dagger, P_{ij}\partial_c F_j\right\rangle + \int_0^T \partial_c o \; dt
\end{aligned}
$$

Our purpose of adjoint formulation is to avoid calculation of $\partial_c x_k$. The adjoint formulation is required for first two terms of $\mathcal{H}$:

- $\left\langle x_i^\dagger, \frac{d}{dt}\partial_c x_i\right\rangle$

$$
\begin{aligned}
\left\langle x_i^\dagger, \frac{d}{dt}\partial_c x_i\right\rangle &= \int_0^T x_i^\dagger \frac{d}{dt}\partial_c x_i \; dt \\
&= x_i^\dagger \partial_c x_i \Big|_0^T - \int_0^T \frac{d}{dt}x_i^\dagger \partial_c x_i \; dt \\
&= x_i^\dagger \partial_c x_i \Big|_0^T - \left\langle \frac{d}{dt}x_i^\dagger, \partial_c x_i\right\rangle
\end{aligned}
$$

- $\left\langle x_i^\dagger, \left[\partial_k P_{ij} \cdot F_j + P_{ij} \cdot \partial_k F_j\right] \partial_c x_k \right\rangle$

$$
\begin{aligned}
&\left\langle x_i^\dagger, \left[\partial_k P_{ij} \cdot F_j + P_{ij} \cdot \partial_k F_j\right] \partial_c x_k \right\rangle \\
&= \int_0^T x_i^\dagger \left[\partial_k P_{ij} \cdot F_j + P_{ij} \cdot \partial_k F_j\right] \partial_c x_k \, dt \\
&= \int_0^T x_k^\dagger \left[\partial_i P_{kj} \cdot F_j + P_{kj} \cdot \partial_i F_j\right] \partial_c x_i \, dt \\
&= \left\langle \left[\partial_i P_{kj} \cdot F_j + P_{kj} \cdot \partial_i F_j\right] x_k^\dagger, \partial_c x_i \right\rangle
\end{aligned}
$$

  where nothing except the index exchange between i and k has been done, which corresponds to the transpose of the matrix $\in \mathbb{R}^N \times \mathbb{R}^N$.

Substituting these two terms, we obtain the dual expression for the sensitivity,

$$
\begin{aligned}
\nabla_\theta \mathcal{O} &= \langle \nabla_x o, \nabla_\theta \mathbf{x} \rangle + \mathcal{H} + \int_0^T \nabla_\theta o \, dt \\
&= \langle \partial_i o, \partial_c x_i \rangle + \left\langle x_i^\dagger, \frac{d}{dt} \partial_c x_i \right\rangle - \left\langle x_i^\dagger, \left[\partial_k P_{ij} \cdot F_j + P_{ij} \cdot \partial_k F_j\right] \partial_c x_k \right\rangle \\
&\qquad - \left\langle x_i^\dagger, \partial_c P_{ij} F_j \right\rangle - \left\langle x_i^\dagger, P_{ij} \partial_c F_j \right\rangle + \int_0^T \partial_c o \, dt \\
&= \langle \partial_i o, \partial_c x_i \rangle - \left\langle \frac{d}{dt} x_i^\dagger, \partial_c x_i \right\rangle - \left\langle \left[\partial_i P_{kj} \cdot F_j + P_{kj} \cdot \partial_i F_j\right] x_k^\dagger, \partial_c x_i \right\rangle \\
&\qquad + x_i^\dagger \partial_c x_i \Big|_0^T - \left\langle x_i^\dagger, \partial_c P_{ij} F_j \right\rangle - \left\langle x_i^\dagger, P_{ij} \partial_c F_j \right\rangle + \int_0^T \partial_c o \, dt \\
&= \left\langle -\frac{d}{dt} x_i^\dagger - \left[\partial_i P_{kj} \cdot F_j + P_{kj} \cdot \partial_i F_j\right] x_k^\dagger + \partial_i o, \partial_c x_i \right\rangle \\
&\qquad + x_i^\dagger \partial_c x_i \Big|_0^T - \left\langle x_i^\dagger, \partial_c P_{ij} F_j \right\rangle - \left\langle x_i^\dagger, P_{ij} \partial_c F_j \right\rangle + \int_0^T \partial_c o \, dt
\end{aligned}
$$

First term in the equation above is now the new constraint functional, i.e. the adjoint equation, and the other three terms are equivalent to the sensitivity $\nabla_\theta \mathcal{O}$. As long as the adjoint equation equals to 0, $\partial_c x_i$ needs not be computed. Regarding the second term $x_i^\dagger \partial_c x_i$, we don't really have the information about $\partial_c \mathbf{x}$ at T, nor use the final condition of $\mathbf{x}$ as a control parameter. This provides us with the final condition for $\mathbf{x}^\dagger$,

$$
\mathbf{x}^\dagger = 0 \qquad \text{at } t = T
$$

Therefore, the sensitivity $\nabla_\theta \mathcal{O}$ can be computed by solving the following adjoint system:

$$
\begin{aligned}
\frac{d}{dt} x_i^\dagger &= -\left[\partial_i P_{kj} \cdot F_j + P_{kj} \cdot \partial_i F_j\right] x_k^\dagger + \partial_i o \\
\partial_k P_{ij} &= -\partial_k V_{il} U_{lj} - V_{il} \partial_k U_{lj} \\
\mathbf{x}^\dagger &= 0 \qquad \text{at } t = T
\end{aligned}
$$

$$
\begin{aligned}
\partial_c \mathcal{O} &= -x_i^\dagger \partial_c x_i \Big|_{t=0} - \left\langle x_i^\dagger, \partial_c P_{ij} F_j \right\rangle - \left\langle x_i^\dagger, P_{ij} \partial_c F_j \right\rangle + \int_0^T \partial_c o \, dt \\
\partial_c P_{ij} &= -\partial_c V_{il} U_{lj} - V_{il} \partial_c U_{lj}
\end{aligned}
$$

Summary

- Governing equation

$$
\begin{aligned}
\frac{d\mathbf{x}}{dt} &= \mathbf{P}(\mathbf{x};\theta)\cdot\mathbf{F}(\mathbf{x};\theta)\\
\mathbf{P}(\mathbf{x};\theta) &= \mathbf{I} - \mathbf{V}_f\mathbf{U}_f\\
\mathbf{J} = \nabla_x\mathbf{F} &= \begin{bmatrix}\mathbf{V}_s & \mathbf{V}_f\end{bmatrix}\begin{bmatrix}\Lambda_{ss} & 0\\ 0 & \Lambda_{ff}\end{bmatrix}\begin{bmatrix}\mathbf{U}_s\\ \mathbf{U}_f\end{bmatrix}
\end{aligned}
\tag{6}
$$

- Observable and its sensitivity

$$
\begin{aligned}
\mathcal{O}[\mathbf{x};\theta] &= \int_0^T o(\mathbf{x};\theta)dt\\
\nabla_\theta\mathcal{O} &= \langle\nabla_x o, \nabla_\theta\mathbf{x}\rangle + \int_0^T \nabla_\theta o\, dt
\end{aligned}
\tag{7}
$$

- Adjoint equation

$$
\begin{aligned}
\frac{d}{dt}x_i^\dagger &= -\left[\partial_i P_{kj}\cdot F_j + P_{kj}\cdot\partial_i F_j\right]x_k^\dagger + \partial_i o\\
\partial_k P_{ij} &= -\partial_k V_{il}U_{lj} - V_{il}\partial_k U_{lj}\\
\mathbf{x}^\dagger &= 0 \qquad\qquad \text{at } t = T
\end{aligned}
\tag{8}
$$

  - i, j, k: the index on $\mathbb{R}^N$ space
  - c: the index on control space $\mathbb{R}^{N_c}$
  - l: the index on fast manifold space $\mathbb{R}^{N_f}$

- Adjoint-based sensitivity

$$
\begin{aligned}
\partial_c\mathcal{O} &= -x_i^\dagger\partial_c x_i\Big|_{t=0} - \left\langle x_i^\dagger, \partial_c P_{ij}F_j\right\rangle - \left\langle x_i^\dagger, P_{ij}\partial_c F_j\right\rangle + \int_0^T \partial_c o\, dt\\
\partial_c P_{ij} &= -\partial_c V_{il}U_{lj} - V_{il}\partial_c U_{lj}
\end{aligned}
\tag{9}
$$

Comments on solving the adjoint equation

In solving the adjoint equation (8), we may first think of the same manifold-reduction as we do for the governing equation. That is, we can decompose $\mathbf{x}^\dagger$ into slow and fast manifolds according to the matrix

$$
-\left[\partial_i P_{kj}\cdot F_j + P_{kj}\cdot\partial_i F_j\right],
$$

and reduce the system onto slow manifolds.
Although, this procedure may not be necessary and rather redundant. On this matter you should have a better knowledge, but I guess manifold reduction is usually required for a nonlinear system which is very stiff until it reaches the slow manifolds. On the other hand, the adjoint system (8), although time-variant, it is basically a linear system. Of course I cannot guarantee that its stiffness is always smaller than that of the original nonlinear system. But I hope a direct time-integration for (8) may be possible in an efficient way, but simpler than just following the same decomposition for nonlinear systems.