1 Jacobian-vector products

Consider the parameterized ODE initial value problem

$$\dot{y} = f(t, y, a), \qquad y(0, a) = y_0(a), \tag{1}$$

by which we mean

$$\partial_0 y(t, a)[1] = f(t, y(t, a), a), \qquad y(0, a) = y_0(a), \tag{2}$$

for all t and a in some domains. We want to understand how the solution to the ODE changes (e.g. at particular values of t) for small perturbations of a. That is, we want to be able to compute the Jacobian-vector product

$$(a, v) \mapsto \partial_1 y(t, a)[v] \tag{3}$$

at any particular values of t and a, where v can be interpreted as a small perturbation to the value of a.

Since the ODE holds true for all values of a (or at least those close to a particular a_0 in which we are interested), we can view both sides as functions of a, and assuming differentiability we can differentiate both sides with respect to a to find a new equation that must be satisfied:

$$\partial_1(\,(t, a) \mapsto \partial_0 y(t, a)[1]\,) = \partial_2 f(t, y(t, a), a) + \partial_1 f(t, y(t, a), a) \circ \partial_1 y(t, a). \tag{4}$$

Applying both sides to a particular perturbation vector v and using the fact that partial derivaives commute, we have

$$\partial_0(\,(t, a) \mapsto \partial_1 y(t, a)[v]\,)[1] = \partial_2 f(t, y(t, a), a)[v] + \partial_1 f(t, y(t, a), a)[\,\partial_1 y(t, a)[v]\,].$$

We can identify $z(t, a) \triangleq \partial_1 y(t, a)[v]$ as a new state vector to write a joint ODE system

$$\begin{bmatrix} \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} f(t, y, a) \\ g(t, y, z, a) \end{bmatrix}, \qquad \begin{bmatrix} y(0, a) \\ z(0, a) \end{bmatrix} = \begin{bmatrix} y_0(a) \\ \partial y_0(a)[v] \end{bmatrix}, \tag{5}$$

$$g(t, y, z, a) = \partial_1 f(t, y, a)[z] + \partial_2 f(t, y, a)[v]. \tag{6}$$

Notice that the dynamics on the z component are linear/affine in z (and v!).

2 Vector-Jacobian products

Consider the parameterized linear/affine ODE IVP

$$\dot{z}(t) = A(t)z(t) + B(t)v, \qquad z(0) = Cv. \tag{7}$$

The implicit mapping $\mathcal{T}_1 : v \mapsto z$ is linear, and so for any linear functional on solution functions $\mathcal{T}_2 : z \mapsto \mathbb{R}$ there is linear function on perturbations v defined by $\mathcal{T}_2 \circ \mathcal{T}_1 : v \mapsto \mathbb{R}$. Given such a linear functional $\mathcal{T}_2$, we wish to find an explicit representer vector for $\mathcal{T}_2 \circ \mathcal{T}_1$.

Concretely, consider linear functionals of the form

$$\mathcal{T}_2[z] = w_T^{\mathsf{T}} z(T) + \int_0^T w(t)^{\mathsf{T}} z(t)\,\mathrm{d}t. \tag{8}$$

We wish to find a function $\lambda(t)$ such that

$$w_T^{\mathsf{T}} z(T) + \int_0^T w(t)^{\mathsf{T}} z(t)\,\mathrm{d}t = \lambda(0)^{\mathsf{T}} Cv + \int_0^T \lambda(t)^{\mathsf{T}} B(t)[v]\,\mathrm{d}t, \tag{9}$$

for all v, when z is a solution to (7).

Consider first the special case when $B \equiv 0$, so that we have the ODE

$$\dot{z}(t) = A(t)z(t), \qquad z(0) = Cv. \tag{10}$$

Moreover consider the special case of a weighted evaluation functional

$$\mathcal{T}_2[z] = w^{\mathsf{T}} z(T). \tag{11}$$

We wish to find a representer vector λ such that

$$\lambda^{\mathsf{T}} z(0) = w^{\mathsf{T}} z(T). \tag{12}$$

Since the particular time $t = 0$ is arbitrary, a more general problem would be to find a function $\lambda(t)$ such that

$$\lambda(t)^{\mathsf{T}} z(t) = w^{\mathsf{T}} z(T). \tag{13}$$

That is, we can fix $\lambda(T) = w$ and ensure that the value of $\lambda(t)^{\mathsf{T}} z(t)$ does not change with time:

$$0 = \partial(\, t \mapsto \lambda(t)^{\mathsf{T}} z(t)\,) = \dot{\lambda}(t)^{\mathsf{T}} z(t) + \lambda(t)^{\mathsf{T}} \dot{z}(t) \tag{14}$$

$$= \dot{\lambda}(t)^{\mathsf{T}} z(t) + \lambda(t)^{\mathsf{T}} A(t) z(t), \tag{15}$$

where on the last line we have used the ODE (10). To satisfy this equation for all t and nonzero solutions $z(t)$, we can choose

$$\dot{\lambda}(t) = -A(t)^{\mathsf{T}} \lambda(t), \qquad \lambda(T) = w. \tag{16}$$

This gives us a means of computing a representer for the linear functional $\mathcal{T}_2$ by solving another ODE IVP, integrating backward in time from $t = T$ to $t = 0$ to compute $\lambda(0)$. (Notice that by linearity we can handle a linear functional that is a linear combination of such weighted evaluation functionals, say at times $0 < T_1 < T_2$, by pulling back the functional at time $t = T_2$ to a representer at time $t = T_1$ and summing before pulling back the sum to $t = 0$.)