GRADIENTS AND NONLINEAR LEAST SQUARES

Steve Cox, CAAM at Rice

1. Introduction

We consider the dynamical system

$$y'(t) = F(t, y(t), p), \qquad y(0) = y_0 \tag{1.1}$$

where F is a mapping from $\mathbf{R} \times \mathbf{R}^N \times \mathbf{R}^M$ to $\mathbf{R}^N$.

2. Full Measurement

We suppose y is known, call it $y^\sharp$ on some interval, $[0,T]$, and we ask how to recover the parameter set, $p^\sharp$, that gave rise to this $y^\sharp$. We do this by solving

$$\min_{p\in\mathbf{R}^M} J(p), \qquad J(p) \equiv \int_0^T \|y(t;p) - y^\sharp(t)\|^2\,dt \tag{2.1}$$

where $y(t;p)$ is the solution to (1.1) for the current choice of p. Or interest here is in effective ways of computing the gradient of the objective function, J, in (2.1).

We consider the Lagrangian

$$L(y,z,p) \equiv \int_0^T \|y(t) - y^\sharp(t)\|^2\,dt + \int_0^T z(t)\cdot(y'(t) - F(t,y(t),p))\,dt \tag{2.2}$$

and trust in the fact that

$$J(p) = \min_y \max_z \; L(y,z,p)$$

and so look for critical points, in y and z, of L. To begin, the directional derivative

$$L_y(y,z,p)\cdot\tilde{y} = 2\int_0^T (y(t) - y^\sharp(t))\cdot\tilde{y}(t)\,dt + \int_0^T z(t)\cdot(\tilde{y}' - F_y(t,y(t),p)\tilde{y}(t))\,dt$$

after integration by parts brings

$$L_y(y,z,p)\cdot\tilde{y} = 2\int_0^T (y(t)-y^\sharp(t))\cdot\tilde{y}(t)\,dt + z(T)\cdot\tilde{y}(T) - \int_0^T (z'(t)+F_y^*(t,y(t),p)z(t))\cdot\tilde{y}(t))\,dt$$

As this must vanish for each $\tilde{y}$ it follows that z must obey the terminal value problem

$$z'(t) + F_y^*(t,y(t),p)z(t) = 2(y(t) - y^\sharp(t)), \qquad z(T) = 0. \tag{2.3}$$

With y and z in hand one may proceed to compute

$$J_p(p)\cdot\tilde{p} = L_p(y,z,p)\cdot\tilde{p} = -\int_0^T z(t)\cdot F_p(t,y(t),p)\tilde{p}\,dt$$

Let us apply this in the case that

$$F(t,y,p) = \begin{pmatrix} p_1 y_1 + p_2 y_1 y_2 \\ p_3 y_1 + p_4 y_2 \end{pmatrix}$$

One easily finds

$$F_y(t,y(t),p) = \begin{pmatrix} p_1 + p_2 y_2 & p_2 y_1 \\ p_3 & p_4 \end{pmatrix} \quad \text{and} \quad F_p(t,y(t),p) = \begin{pmatrix} y_1 & y_1 y_2 & 0 & 0 \\ 0 & 0 & y_1 & y_2 \end{pmatrix}$$

3. **Spiky Data**

In situtations with rapid transients it can be useful to use a J of the form

$$J(p) \equiv \int_0^T \|Y(t;p) - Y^\sharp(t)\|^2\, dt$$

where Y is the running integral of y, i.e.,

$$Y(t;p) \equiv \int_0^t y(s;p)\, ds.$$

Now the leading term in the y derivative of the associated Lagrangian is

$$2\int_0^T (\mu(T) - \mu(t)) \cdot \tilde{y}(t)\, dt$$

where

$$\mu(t) \equiv \int_0^t (t-s)(y(s) - y^\sharp(s))\, ds$$

4. **Pointwise Measurement**

In the case that we only measure the state at the specific times, $\{t_k\}_{k=1}^K$, our objective function takes the form

$$J(p) \equiv \sum_{k=1}^K \|y(t_k;p) - y^\sharp(t_k)\|^2 = \int_0^T \sum_{k=1}^K \|y(t;p) - y^\sharp(t)\|^2 \delta(t - t_k)\, dt$$

Now all that changes is the righthand side of the adjoint equation. Namely, we must solve

$$z'(t) + F_y^*(t,y(t),p)z(t) = 2(y(t) - y^\sharp(t)) \sum_{k=1}^K \delta(t - t_k), \qquad z(T) = 0. \tag{4.1}$$

This is, in fact, a sequence of simple problems between the measurement times. First,

$$
\begin{aligned}
z'(t;K) + F_y^*(t, y(t), p) z(t;K) = 0, \quad t_{K-1} < t < t_K, \\
z(t_K;K) = 2(y^\sharp(t_K) - y(t_K))
\end{aligned}
$$

next

$$
\begin{aligned}
z'(t;K-1) + F_y^*(t, y(t), p) z(t;K-1) = 0, \quad t_{K-2} < t < t_{K-1}, \\
z(t_{K-1};K-1) = z(t_{K-1};K) - 2(y(t_{K-1}) - y^\sharp(t_{K-1}))
\end{aligned}
$$

and in this fashion until

$$
\begin{aligned}
z'(t;1) + F_y^*(t, y(t), p) z(t;1) = 0, \quad 0 < t < t_1, \\
z(t_1;1) = z(t_1;2) - 2(y(t_1) - y^\sharp(t_1))
\end{aligned}
$$