Lecture 39: Nonlinear Parameter Estimation.

(See Smith, Uncertainty Quantification, Section 7.3)

Thus far we have only considered the linear model

$$Y = Xq_0 + \varepsilon,$$

i.e., our design Xq only involves $q_1, q_2, \ldots, q_n$ in a linear fashion. Many interesting models give nonlinear dependence of the observations on the quantities of interest. This can be true even when the underlying dynamical system is linear. For example, consider this example from Smith's book:

A damped mass-spring system can be modeled by the second-order linear differential equation

$$z''(t) + Cz'(t) + Kz(t) = 0$$

with initial conditions

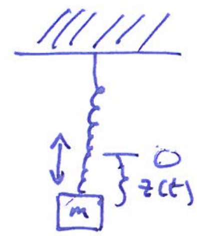
$$z(0) = 2, \quad z'(0) = -C$$ } Chosen to give a nice form for the solution.

If $C=0$, $K>0$, this system is undamped: it will vibrate forever. When $C>0$, the system will lose energy as it vibrates.

If you build a simple mass-spring system, you will see its vibration damp out in time, due to damping from the air and internal mechanical forces. It is difficult to measure C directly. (We have better hope of estimating K from the mass of the object and the stiffness of the spring.)

Even though the differential equation is linear, the solution

$$z(t) = 2e^{-Ct/2} \cos\left(\sqrt{K - C^2/4}\, t\right)$$



shows that the displacement $z(t)$ depends nonlinearly on C.

This is an instance of the general nonlinear model

$$Y = f(q_0) + \varepsilon \qquad \begin{array}{l} \mathbb{E}(\varepsilon) = 0 \\ Var(\varepsilon_i) = \sigma_0^2,\ Cov(\varepsilon_i, \varepsilon_j) = 0 \end{array}$$

(Our linear model fits into this framework too! $f(q_0) = Xq_0$.)

To estimate q_0, we again seek to minimize the least squares error functional:

Pick $\hat{q}$ such that

$$\|Y - f(\hat{q})\|^2 = \min_{q \in Q} \|Y - f(q)\|^2 \qquad (*)$$

$$= \min_{q \in Q} J(q).$$

Here Q is the set of admissible values for the parameters q. For example, you can use Q to enforce nonnegative values of q, if you like.

In the case $Q = \mathbb{R}^p$, we can solve $(*)$ using the MATLAB command "Lsqnonlin", which is part of the Optimization Toolbox. This can be <u>much trickier</u> than solving the linear least squares problem in Lecture 36. In particular, the objective function $J(q)$ can have <u>local minima</u>, so Lsqnonlin can converge to an <u>incorrect solution</u>.

The performance of Lsqnonlin depends on giving an accurate <u>initial guess</u> to $\hat{q}$. For hard problems, you might need to do some preliminary work to get good initial guesses.

Given the estimate $\hat{q}$ for the QoI's, we would still like to estimate $\mathrm{Var}(\hat{q})$ and σ_0^2.

To do so, we <u>linearize</u> the nonlinear problem in the neighborhood of $\hat{q}$. The Taylor series in higher dimensions takes the form

$$f(q) = f(\hat{q}) + X(q-\hat{q}) + O(\|q-\hat{q}\|^2)$$

where X is the Jacobian of f at $\hat{q}$:

$$X_{j,k} = (X(\hat{q}))_{j,k} = \frac{\partial f_j}{\partial q_k}(\hat{q}).$$

<u>Examples</u>

- If we have a linear model, $Y = Xq_0 + \varepsilon$, then $f(q) = Xq$, so $f(q) = f(\hat{q}) + X(q-\hat{q}) + O(\|q-\hat{q}\|^2)$ reduces to $Xq = X\hat{q} + X(q-\hat{q}) + 0$, so this approach is consistent with our earlier work.

- For the damped mass-spring system, where we seek C, we have: $q = [C] \in \mathbb{R}^p$ for $p=1$

$$f(q) = z(t;C) = 2e^{-Ct/2}\cos\left(\sqrt{K - C^2/4}\,t\right)$$

$$X = \begin{bmatrix} \frac{\partial z(t_1;C)}{\partial C} \\ \vdots \\ \frac{\partial z(t_n;C)}{\partial C} \end{bmatrix} \text{ evaluated at } \hat{q} = \hat{C}. \in \mathbb{R}^{n\times 1}$$

- If we seek both c and K, we have

$$q = \begin{bmatrix} K \\ c \end{bmatrix} \in \mathbb{R}^2 \quad \text{for } p = 2$$

$$f(q) = z(t; K, c) = 2e^{-ct/2} \cos\left(\sqrt{K - c^2/4}\, t\right)$$

$$X = \begin{bmatrix} \frac{\partial z(t_1; K, c)}{\partial K} & \frac{\partial z(t_1; K, c)}{\partial c} \\ \vdots & \\ \frac{\partial z(t_n; K, c)}{\partial K} & \frac{\partial z(t_n; K, c)}{\partial c} \end{bmatrix} \in \mathbb{R}^{n \times 2} \quad \text{evaluated at } \hat{q} = \begin{bmatrix} \hat{K} \\ \hat{c} \end{bmatrix}$$

- From this <u>linearized</u> model, we have

$$\mathrm{Var}(\hat{q}) = \sigma_0^2 (X^T X)^{-1} \qquad \left(X = \text{Jacobian evaluated at } \hat{q}. \right)$$

$$\hat{\sigma}_0^2 = \frac{\|\hat{R}\|^2}{n - p} \quad \text{for } \hat{R} = Y - f(\hat{q})$$

In practical cases, we sometimes have p very large, making these parameter estimation problems very difficult: a major modern computational science challenge!

For examples with $q = [c]$ and $q = \begin{bmatrix} K \\ c \end{bmatrix}$ described above, see est_shm.m on the class website.