Sensitivity to Model Parameters

C. David Levermore
Department of Mathematics *and*
Institute for Physical Science and Technology
University of Maryland, College Park
lvrmr@math.umd.edu

Math 420: *Mathematical Modeling*
9 March 2011

Outline

1) Introduction
2) Ensemble Sensitivity
3) Linear (Forward) Sensitivity
4) Adjoint (Backward) Sensitivity
5) Example
6) Model Inflation
7) Further Questions

1. Introduction: The Problem

Suppose that for some model the state of the system being modeled is described by a vector unknown U determined by an equation that depends upon a vector of parameters P in the form

$$F(U, P) = 0\,.$$

Here we assume that for every P there exists a unique solution $U(P)$ of this equation that depends upon P in a nice way. This is the notion of the equation being well-posed.

Suppose that the output of our model we care about is given by an objective function $G(U, P)$. The question we want to address is

How sensitive is the response $R(P) = G(U(P), P)$ to changes in the parameters P?

1. Introduction: Why?

The answer to this question might help us

- find out which parameters have to be calibrated more carefully;
- quantify the uncertainty in predictions of the model;
- see how the model might be improved or expanded.

We will illustrate ideas about sensitivity in the nonlinear setting in which the unknown vector U has dimension M, the parameter vector P has dimension $N >> M$, and the objective function $G(U, P)$ takes values in dimension $K << M$. (Often $K = 1$ or some other small integer.)

2. Ensemble Sensitivity

Suppose that our best guess for the values of the parameters in the model is P_o. Let $U_o = U(P_o)$ be the associated solution of our equation. The response of our model is then $R(P_o) = G(U_o, P_o)$. The question is then how $R(P) = G(U(P), P)$ behaves when P is drawn from alternative values "near" P_o?

The brute force approach to this question is to compute $U(P)$ for an ensemble of values P drawn from an N-dimensional hypercube around P_o and then compute the associated $R(p)$.

If we sample each vertex of such a hypercube then we must solve for U exactly 2^N times! OUCH! If we sample each face of such a hypercube then we must solve for U exactly $2N$ times! Much better, but still ouch! How we pick the hypercube is a bit arbitrary.

3. Linearized (Forward) Sensitivity

A less arbitrary approach is to compute the linearized sensitivity of the response $R(P) = G(U(P), P)$. Specifically, we set $U_o = U(P_o)$ and want to compute

$$\partial_P R(P_o) = G_U(U_o, P_o) U_P(P_o) + G_P(U_o, P_o) \, .$$

Here we know $G_U(U_o, P_o)$ and $G_P(U_o, P_o)$, but not $U_P(P_o)$.

By differentiating the model equation for $U(P)$ we see that $U_P(P_o)$ satisfies the linearized model

$$0 = F_U(U_o, P_o) U_P(P_o) + F_P(U_o, P_o) \, .$$

When $N > M$ this approach generally requires inverting $F_U(U_o, P_o)$ by Gaussian elimination. This is no more expensive than solving for U about M times. Because $M << N$, this is much cheaper than the ensemble approach.

4. Adjoint (Backward) Sensitivity

Better still, we find the row-vector J_o that solves the adjoint problem

$$0 = J_o F_U(U_o, P_o) + G_U(U_o, P_o)\,.$$

Then the sensitivity becomes

$$\begin{aligned}\partial_P R(P_o) &= -J_o F_U(U_o, P_o) U_P(P_o) + G_P(U_o, P_o) \\ &= J_o F_P(U_o, P_o) + G_P(U_o, P_o)\,.\end{aligned}$$

Because we know $F_P(U_o, P_o)$ and $G_P(U_o, P_o)$, this is easy to compute once J_o is computed.

The point here is that computing J_o only requires solving K linear systems. The cost of doing this will be roughly K times the cost of solving for U_o. Because $K << M$ this will be much less than the cost of solving the linearized model for $U_P(P_o)$, which will generally be M times the cost of solving for U_o.

5. Example

Suppose our model $u(t,p)$ is governed by a first-order system as

$$\frac{\mathrm{d}u}{\mathrm{d}t} = f(t,u,p)\,, \qquad u(0,p) = u^{\mathsf{in}}(p)\,,$$

where p is a vector of parameters. Suppose our response function has the form $r \cdot u(T,p)$ where r is a row vector and $T > 0$ is a time. Let p_o be our best guess for the values of p.

First solve the above initial-value problem for $u_o(t) = u(t,p_o)$ over the time interval $[0,T]$. Next solve the adjoint initial-value problem

$$-\frac{\mathrm{d}w}{\mathrm{d}t} = w \cdot f_u(t,u_o(t),p_o)\,, \qquad w(T) = r\,.$$

Then the linearized sensitivity of the response is

$$r \cdot u_p(T,p_o) = w(0) \cdot u_p^{\mathsf{in}}(p_o) + \int_0^T w(t) \cdot f_p(t,u_o(t),p_o)\,\mathrm{d}t\,.$$

This formula can be derived by noticing that $v_p(t) = u_p(t, p_o)$ satifies the linearized initial-value problem

$$\frac{\mathrm{d}v_p}{\mathrm{d}t} = f_u(t, u_o(t), p_o)v_p + f_p(t, u_o(t), p_o)\,, \qquad v_p(0) = u_p^{\mathrm{in}}(p_o)\,.$$

It then follows that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(w \cdot v_p) &= \frac{\mathrm{d}w}{\mathrm{d}t} \cdot v_p + w \cdot \frac{\mathrm{d}v_p}{\mathrm{d}t} \\ &= -w \cdot f_u(t, u_o(t), p_o)v_p \\ &\quad + w \cdot \Big(f_u(t, u_o(t), p_o)v_p + f_p(t, u_o(t), p_o)\Big) \\ &= w \cdot f_p(t, u_o(t), p_o)\,.\end{aligned}$$

Our formula is obtained by integrating this relation over $[0, T]$.

6. Model Inflation

Model inflation builds larger models from smaller ones. Here we present an approach that has four ingredients.

1. Imbed your current model in a large family of larger models.

2. Identify a few objective functions that you wish to predict.

3. Use adjoint sensitivity analysis to compute the linear response of your current objective functions to every parameter in the family.

4. Enlarge your current model to capture the parameters to which your objective functions are most sensitive and repeat.

We illustrate the approach in the nonlinear setting in which the family of larger models for the unknown vector U of dimension M has the form

$$F(U, P) = 0\,,$$

where the parameter vector P has dimension $N >> M$. We assume that for every P there exists a unique solution $U(P)$ of this equation. We also assume that when $P = P_o$ this family reduces to our current model.

Let ΔP_o be a diagonal matrix of uncertainties assosiated with P_o. If an entry of P_o is the mean of some data then the corresponding entry of ΔP_o might be the standard deviation of that data. If an entry of P_o is zero in order to turn off or decouple some phenomenon then the corresponding entry of ΔP_o might be your best guess at the expected value of that parameter.

Now let $\triangle R_o$ be the $K \times N$ matrix whose entries are the absolute values of the entries of $\partial_P R(P_o) \triangle P_o$. The entries of this matrix are your best guesses of the uncertainties in the objective functions.

The idea is now to use the matrix $\triangle R_o$ to identify which parameters that are zero in the current model should be turned on so as to enlarge the model. One way to do this is to simply choose those parameters corresponding to the largest entry in each row of $\triangle R_o$. One can also consider some convex combinations of the entries in each row of $\triangle R_o$. One then enlarges the current model and repeats the process.

The model inflation stops when none of the largest uncertainties are due to parameters that are zero. When this happens you can not make better predictions by enlarging the model.

A model should be as simple as possible, but no simpler!

7. Further Questions

Some natural questions arise from the foregoing discussion.

- What can be done for models whose governing equation cannot be linearized other than ensemble sensitivity?
- When is a model be linearly insensitive, but nonlinearly sensitive?
- Do these ideas extend to stochastic models (ones with randomness)?
- How can the family of models assumed in our discussion of model inflation be constructed?