Derivation of Gradient Test

October 28, 2015

Take any twice continuously differentiable function $f(x) : \mathbf{R}^n \to \mathbf{R}$. For two nearby points x_0 and x_1, from the first order taylor expansion we have

$$\begin{aligned} f(x_1) &= f(x_0) + \nabla f(x_0)^T(x_1 - x_0) + \tfrac{1}{2}(x_1 - x_0)^T \nabla^2 f(x_0)(x_1 - x_0) + o\left(\|x_1 - x_0\|^2\right) \\ f(x_0) &= f(x_1) + \nabla f(x_1)^T(x_0 - x_1) + \tfrac{1}{2}(x_0 - x_1)^T \nabla^2 f(x_0)(x_0 - x_1) + o\left(\|x_0 - x_1\|^2\right). \end{aligned}$$

Subtracting these equations gives

$$\begin{aligned} f(x_1) - f(x_0) &= f(x_0) - f(x_1) + \nabla f(x_0)^T(x_1 - x_0) - \nabla f(x_1)^T(x_0 - x_1) + o\left(\|x_1 - x_0\|^2\right) \\ &= -\left(f(x_1) - f(x_0)\right) + \left(\nabla f(x_0) + \nabla f(x_1)\right)^T (x_1 - x_0) + o\left(\|x_1 - x_0\|^2\right). \end{aligned}$$

Adding the expression $f(x_1) - f(x_0)$ to both sides and rearranging gives

$$1 + \frac{o\left(\|x_1 - x_0\|^2\right)}{2 * \left(f(x_1) - f(x_0)\right)} = \frac{1}{2}\frac{\left(\nabla f(x_0) + \nabla f(x_1)\right)^T (x_1 - x_0)}{f(x_1) - f(x_0)}.$$

The expression $o\left(\|x_1 - x_0\|^2\right)$ represents any quantity that satisfies

$$\lim_{\|x_1 - x_0\| \to 0} \frac{o(\|x_1 - x_0\|^2)}{\|x_1 - x_0\|^2} = 0,$$

in other words, a quantity that goes to 0 faster than $\|x_1 - x_0\|^2$. Thus, for a small enough quantity $\|x_1 - x_0\|$, we should have

$$1 \approx \frac{1}{2}\frac{\left(\nabla f(x_0) + \nabla f(x_1)\right)^T (x_1 - x_0)}{f(x_1) - f(x_0)}.$$

unless $f(x_1) - f(x_0) \approx \|x_1 - x_0\|^{2+\epsilon}$ where $\epsilon > 1$. In an application setting this may indicate that the model function f is insufficiently sensitive to the parameters of interest.

A reasonably robust test to check if the gradient is computed correctly is to verify the following:

$$\begin{aligned} \left|\frac{\|x_1 - x_0\|^2}{2(f(x_1) - f(x_0))}\right| &\leq c_1 \\ \left|1 - \frac{1}{2}\frac{\left(\nabla f(x_0) + \nabla f(x_1)\right)^T (x_1 - x_0)}{f(x_1) - f(x_0)}\right| &\leq c_2. \end{aligned}$$

The values c_1 and c_2 may depend on the application, but we can try taking $c_1 = c_2 = \|x_1 - x_0\| = 10^{-2}$.

Supposing now that x_0 is a true model, f is a misfit function so $f(x_0) = 0$ and $\nabla f(x_0) = 0$, and Δx is a perturbation added to x_0. The test then becomes

$$\begin{aligned} \frac{\|\Delta x\|^2}{2f(x_1)} &\leq c_1 \\ \left\| 1 - \frac{1}{2}\frac{\nabla f(x_1)^T(\Delta x)}{f(x_1)} \right\| &\leq c_2. \end{aligned}$$

1 Other methods

complex method for derivatives. Add an imaginary amount of a quantity, use Matlab functions to compute the analytic continuation of the model over the complex plane, and dividing the imaginary part of the total level by h. Analogous to computing the derivative of the concentration by a complex-step derivative approximation [4]. [4] is J.R.R.A. Martins, P. Sturdza and J. J. Alonso, The complex-step derivative approximation, ACM TOMS 29:3, 2003.