Evaluation of Functions, Gradients and Jacobians

Louis B. Rall
Department of Mathematics
University of Wisconsin-Madison
Madison, Wisconsin 53706, U.S.A.
rall@math.wisc.edu

A function $f : D \subset \mathbf{R}^d \to \mathbf{R}$ is ordinarily evaluated by constructing a sequence $\{x_1, x_2, \ldots, x_n\}$, where each x_k is obtained by evaluating an expression of the form $x_k = f_k(x_1, \ldots, x_{k-1})$, and $x_n = f(x_1, \ldots, x_d)$. In realization of such algorithms by computer programs, the expressions f_k can be limited to assignments, arithmetic operations, and functions already programmed or built into the hardware, such as those belonging to a standard library of mathematical functions. The number of steps n of the algorithm to evaluate the function may depend on the given values of $x_1, \ldots, x_d$, but this dependence will be suppressed for simplicity of notation.

Instead of considering the evaluation process as a mapping from a point in $\mathbf{R}^d$ to a point in $\mathbf{R}$, it will be viewed as a transformation in $\mathbf{R}^n$ of the form $\mathbf{x} = \mathbf{F}(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, that is, as a fixed point problem in $\mathbf{R}^n$ or the equivalent equation $\mathbf{G}(\mathbf{x}) \equiv \mathbf{x} - \mathbf{F}(\mathbf{x}) = 0$. These formulations give the possibility of improvement of accuracy of the function evaluation and validation of the computed result.

In particular, if the expressions f_k are differentiable, then the Jacobian $J = \mathbf{F}'(\mathbf{x}) = (\partial x_i / \partial x_j)$ exists and is strictly lower triangular. One has $\mathbf{G}'(\mathbf{x}) = I - J$ and $\mathbf{G}'(\mathbf{x})^{-1} = (I - J)^{-1} = I + J + \cdots + J^m$ for some $m \leq n$. Thus, Newton's method can be applied to $\mathbf{G}(\mathbf{x}) = 0$ for improvement of computed values and their validation by interval inclusion.

Another use of the matrix J is the computation of the gradient ∇f, a process commonly referred to as automatic differentiation. The forward mode consists of computing the right eigenvectors of J by the power method, the reverse mode yields a left eigenvector, also by the power method. In either case, accurate matrix-vector multiplication with the aid of a long accumulator and interval validation of results are applicable.

The above results apply immediately to evaluation of functions $f : D \subset \mathbf{R}^\mathbf{p} \to \mathbf{R}^q$ to obtain the corresponding Jacobian of the transformation. In the case $p = q$, the function being computed may be an inverse function, that is, one uses the computer to solve the equation $f(x) = y$ for $x = f^{-1}(y) = g(y)$. Since the program will contain a subroutine for $f(x)$, it is not necessary of find the Jacobian of $g(y)$ by differentiation of the entire routine. Once a satisfactory value of x has been computed, one has $g'(y) = [f'(x)]^{-1}$, so only the subroutine for $f(x)$ has to be differentiated. In this sense, differentiation and inversion commute because of linearity.