HUMBOLDT–UNIVERSITÄT ZU BERLIN
Mathematisch-Naturwissenschaftliche Fakultät II
Institut für Mathematik
Prof. PhD. Andreas Griewank
Kshitij Kulshreshtha

Humboldt-Universität zu Berlin, Institut für Mathematik, Unter den Linden 6, D-10099 Berlin

Introduction to Algorithmic Differentiation

Wintersemester 2004/05

Practice Exercise 2

To be finished by November 10, 2004

Exercise 2.1 Consider an evaluation procedure that contains no additions and subtractions, so that none of the elemental partials $c_{ij} = \partial\phi_i/\partial v_j$ with $j \prec i$ has one of the special values in $\{-1, 0, +1\}$. Moreover, assume that there are no dead ends in that all intermediates v_k with $1 \leq k \leq l - m$ depend on at least one independent variable x_j and impact at least one dependent variable y_i. Finally, assume that no dependent is directly calculated from an independent, so that none of the c_{ij} itself is an element of the Jacobian. Then each value c_{ij} affects the value of the Jacobian element $\partial y_i/\partial x_j$ and must therefore enter somehow into any procedure for evaluating $F'(x)$ at least once. Let the gradient $\nabla f(x) = \overline{F}(x, 1)$ of a scalar-valued function $f = F$ with $m = 1$ be calculated by the adjoint procedure with $\overline{y}_1 = 1$. Then each elemental partial c_{ij} enters exactly once, namely, as a factor in a multiplication. In the incremental form each one of these multiplications is followed by an addition.

a. Conclude that the number of arithmetic operations used to accumulate ∇f from the elemental partials in the reverse mode cannot exceed the minimal number needed to calculate ∇f by a factor of more than 4.
   For some $d \gg 1$ and positive constants w_i, z_i, b_i for $i = 1 \ldots d$, consider the function

$$f(x) = \ln \left| \sum_{i=1}^{d} w_i(\exp(z_i * \sin(x)) - b_i)^2 \right| .$$

   We might think of $\sin(x)$ as a parameter restricted to $[-1, 1]$ that should be chosen to minimise the discrepancies between the values $\exp(z_i * \sin(x))$ and the b_i as measured by the logarithm of a suitably weighted sum. To optimise such an exponential fit, one obviously needs the scalar-valued derivative $f'(x)$. To satisfy the special assumptions made above, let us consider the $\psi_i(u) \equiv (u - b_i)^2$ and the weighted sum $\psi_0(v) = \sum_{i=1}^{d} w_i v_i$ as elemental functions.

b. Write a procedure for evaluating f using elementaries ψ_i for $i = 0, 1, 2, \ldots, d$ and draw the computational graph.

c. Derive the corresponding adjoint procedure for accumulating ∇f from the elemental partials. Count the number of multiplications and additions used in the incremental form. Verify that the number of multiplications corresponds exactly to the number of edges, which is $4d + \mathcal{O}(1)$.

d. Rewrite the formula for $f(x)$ using ψ_i, and differentiate it by hand. Show that by keeping the common factors $\cos(x)$ and $1/\psi_0$ out of the internal sum, the number of multiplications needed for the calculation can be kept to $3d + \mathcal{O}(1)$.

e. Show that the tangent procedure for calculating $f'(x) = \dot{f}(x, \dot{x})$ with $\dot{x} = 1$ involves exactly as many multiplications as the reverse mode and is therefore also less economical than the procedure derived by hand in part d.

Exercise 2.2 Consider the function $f(X) \equiv \ln|\det(X)|$, where $X \in \mathbb{R}^{\hat{n}\times\hat{n}}$ is a square matrix whose $n \equiv \hat{n} \times \hat{n}$ entries X_{ij} are considered as independent variables. This function $f(X)$ has been used as a penalty term for maintaining positive definiteness of X when it is restricted to being symmetric.

a. Show that, provided $\det(X) \neq 0$ so that $f(X)$ can be evaluated, the matrix $G \in \mathbb{R}^{\hat{n}\times\hat{n}}$ of corresponding gradient components $G_{ij} = G_{ij}(X) \equiv \partial f(X)/\partial X_{ij}$ is the transpose of the inverse X^{-1}. Hence the cheap gradient principle guarantees that any finite algorithm for evaluating the determinant of a general square matrix can be transformed into an algorithm for computing the inverse with essentially the same temporal complexity.

b. Examine whether the last assertion is still true when X is restricted to having a certain sparsity pattern (for example, tridiagonal). Explain your conclusion.

c. Assume that $f(X)$ can be evaluated on some neighbourhood of a particular matrix by performing LU factorisation without pivoting. Write this algorithm as an evaluation procedure that overwrites the original matrix elements with intermediate quantities and finally with the non-trivial entries of the two triangular factors. Derive the corresponding adjoint procedure with tape and compare it to standard methods for computing matrix inverses.

d. For a constant matrix $A \in \mathbb{R}^{n\times n}$ and a vector $b \in \mathbb{R}^n$, determine the gradient of

$$f(x) \equiv \ln|\det(A + bx^T)| \quad \text{at} \quad x = 0\,.$$

e. Write an evaluation procedure that evaluates $f(x)$ for $n = 2$ by first computing the entries $A + bx^T$, then applying the explicit determinant formula for 2×2 matrices and finally taking the logarithm of the determinant's modulus. Show that the resulting adjoint procedure for evaluating ∇f is not optimal.

f. Suppose that $A + bx^T$ has for all x near the origin n simple non-zero real eigenvalues $\lambda_j(x)$ so that $f(x) = \sum_j \ln|\lambda_j(x)|$. Determine $\nabla_x \lambda_j$ at $x = 0$ by implicit differentiation of the identities $(A+bx^T)v_j = \lambda_j v_j$ and $(A+bx^T)^T w_j = \lambda_j w_j$. Here v_j, $w_j \in \mathbb{R}^n$ denote the corresponding left and right eigenvectors normalised such that $v_j^T w_j = 1$ for $j = 1, 2, \ldots, n$. Write an evaluation procedure for $f(x)$ treating the λ_j as elemental functions. Derive and interpret the corresponding adjoint procedure for evaluating $\nabla f(0)$.

Phone: (030) 2093-5414 Faxphone: (030) 2093-5848
e-mail: griewank@math.hu-berlin.de, kshitij@math.hu-berlin.de
WWW: http://www.mathematik.hu-berlin.de/∼gaggle/AD/ADuebung.html